(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
17 July 2003 (17.07.2003)

PCT

(10) International Publication Number
WO 03/058198 A2

(51) International Patent Classification[7]: G01N

(21) International Application Number: PCT/US02/41614

(22) International Filing Date:
23 December 2002 (23.12.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/343,907 27 December 2001 (27.12.2001) US

(71) Applicant *(for all designated States except US)*: **CIPHERGEN BIOSYSTEMS, INC.** [US/US]; 6611 Dumbarton Circle, Fremont, CA 94555 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **YIP, Tai-Tung** [US/US]; 1532 Aster Court, Cupertino, CA 95014 (US).

YIP, Christine [US/US]; 1532 Aster Court, Cupertino, CA 95014 (US).

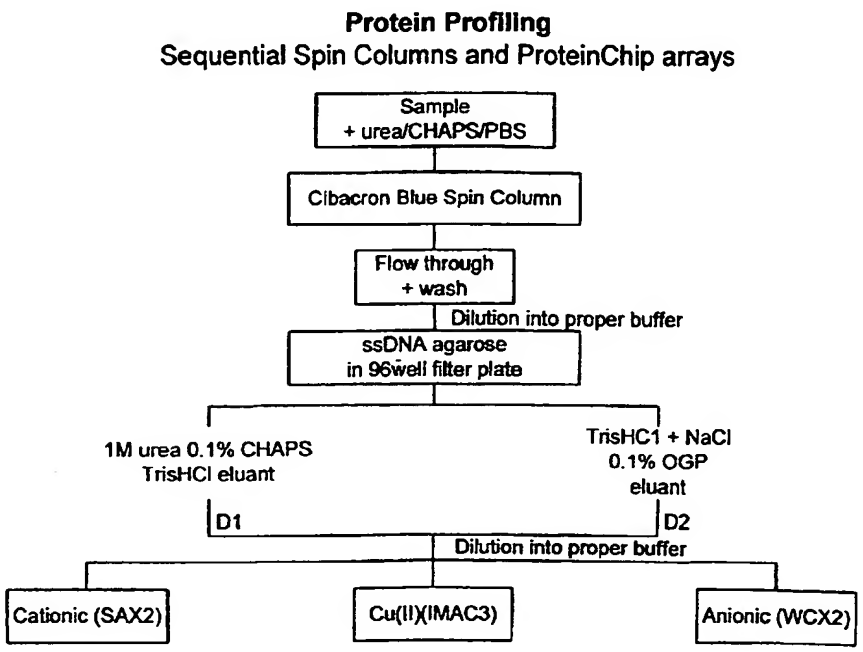
(74) Agents: **WEBER, Kenneth, A.** et al.; Townsend and Townsend and Crew LLP, Two Embarcadero Center, Eighth Floor, San Francisco, CA 94111 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: HUMAN BREAST CANCER BIOMARKERS



(57) Abstract: The invention provides markers, methods and kits that can be used as an aid for breast cancer diagnosis using markers that are differentially present in the samples of breast cancer patients and a control (*e.g.*, women in whom breast cancer is undetectable).

WO 03/058198 A2

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)*

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for all designations*

— *of inventorship (Rule 4.17(iv)) for US only*

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

HUMAN BREAST CANCER BIOMARKERS

BACKGROUND OF THE INVENTION

Breast cancer is the most common fatal malignancy in women. About 15% of all women will be diagnosed with breast cancer during their lifetime. In the U.S., breast cancer is the third leading cause of death in women. Despite recent progress in early detection, as well as improved treatment, the mortality rate remains unchanged. Early diagnosis is the key to surviving breast cancer.

Typically, the detection of breast cancer involves an exam by a physician, a mammogram, and either a needle aspiration or biopsy. When the breast tissue forming the lump is removed, the tissue is examined for possible cancer cells.

A few protein markers have been used to aid the diagnosis and/or prognosis of breast cancer. For example, CA 15-3, a glycoprotein found in the breast, has been used as a marker to monitor recurrence of breast cancer and responses to various treatments. However, the CA 15-3 glycoprotein level is only informative for evaluating patients' condition before and 30 days after the surgical resection of the tumor. Other than in these few cases, this marker is not informative. For example, a third of breast cancer patients with metastasis have CA 15-3 concentrations within the normal range. Even after radical resection of the tumor, CA 15-3 exhibits a substantial variation at abnormal concentrations.

Another marker that has been used to diagnose breast cancer are the neu oncogenes (also known as HER-2 and c-erb-2). This gene is overexpressed in the tumor cells of about 40% women with breast cancer. The neu oncogene products can be elevated as much as 30 to 45-fold above normal levels in human breast tumor. However, there are problems associated with using this protein marker as a diagnostic tool for breast cancer. For example, false positives are often observed using this marker. Moreover, in about 50% of patients, the plasma HER-2 levels do not show any response to treatment. Accordingly, the detection of these individual markers alone cannot be used as a diagnostic test as to whether the subject has a breast cancer.

The effectiveness of any diagnostic test depends on its specificity and selectivity. That is, what is the relative ratio of true positive diagnoses, true negative diagnoses, false positive diagnoses and false negative diagnoses? Methods of increasing

the percent of true positive and true negative diagnoses for any condition are desirable medical goals. In the case of breast cancer, the present diagnostic tests are not completely satisfactory, in that they provide significant numbers of false positive and false negative results as described above.

There is some consensus in the medical community that better diagnosis will result from the discovery of more disease markers. Moreover, in an economy-conscious environment in which cost-effective medicine is an overriding concern, physicians treating cancer patients need convenient, efficient methods to rapidly diagnose breast cancer and to evaluate responses to therapy. The present invention meets this and other needs.

SUMMARY OF THE INVENTION

The present invention provides, for the first time, novel protein markers that are differentially present in the samples of breast cancer patients and in the samples of control subjects. The present invention also provides sensitive and quick methods and kits that can be used as an aid for diagnosis of breast cancer by detecting these novel markers. The measurement of these markers, alone or in combination, in patient samples provides information that diagnostician can correlate with a probable diagnosis of breast cancer or a negative diagnosis (*e.g.*, normal or disease-free). All the markers are characterized by molecular weight. The markers can be resolved from other proteins in a sample by using a variety of fractionation techniques, *e.g.*, chromatographic separation coupled with mass spectrometry, or by traditional immunoassays. In preferred embodiments, the method of resolution involves Surface-Enhanced Laser Desorption/Ionization ("SELDI") mass spectrometry, in which the surface of the mass spectrometry probe comprises adsorbents that bind the markers.

A first set of markers is identified from blood serum, and are capable of binding to a cationic adsorbent, an anionic adsorbent or a metal chelate adsorbent and other adsorbents. These markers include Marker Br 1: 4170 ± 8 Da; Marker Br2: 21080 ± 42 Da; Marker Br 3: 9339 ± 19 Da; Marker Br 4: 28308 ± 57 Da; Marker Br5: 28308 ± 57 Da; Marker Br 6: 4148 ± 9 Da; Marker Br 7: 5634 ± 11 Da; Marker Br 8: 6520 ± 13 Da; Marker Br 9: 6955 ± 14 Da; Marker Br 10: 7507 ± 15 Da; Marker Br 11: 9116 ± 16 Da; Marker Br 12: 9453 ± 19 Da; Marker Br 13: 17310 ± 35 Da; Marker Br 14: 89805 ± 449 Da; Marker Br 15: 4256 ± 9 Da; Marker Br 16: 4357 ± 9 Da; Marker Br 17: 4470 ± 9 Da; Marker Br 18: 9292 ± 19 Da; Marker Br 20: 11758 ± 24 Da; Marker Br 21: 11776 ±

24 Da; Marker Br 22: 13913 ± 28 Da; Marker Br 23: 17291 ± 35 Da; Marker Br 24: 17419 ± 35 Da; Marker Br 25: 21103 ± 42 Da; Marker Br 26: 40297 ± 81 Da; Marker Br 27: 4488 ± 9 Da; Marker Br 28: 4647 ± 9 Da; Marker Br 29: 80428 ± 402 Da; Marker Br 30: 11757 ± 24 Da; Marker Br 31: 4487 ± 9 Da; Marker Br 32: 5360 ± 11 Da; Marker Br 33: 11773 ± 24 Da; Marker Br 34: 13977 ± 28 Da; Marker Br 35: 14004 ± 28 Da; Marker Br 36: 51700 ± 259 Da; and Marker Br 37: 153894 ± 769 Da.

While these markers were first identified from the blood serum sample, the sample from which they can be detected is not limited to a blood serum sample. These markers may be detectable in other types of samples, such as nipple aspirate, urine, tears, saliva, *etc.* Moreover, although the first and second set of markers were discovered using a hydrophilic adsorbent and a metal chelate adsorbent, respectively, the markers are capable of binding other types of adsorbents as described below. Accordingly, embodiments of the invention are not limited to the use of hydrophilic adsorbents and metal chelate adsorbents.

While the absolute identity of these markers is not yet known, such knowledge is not necessary to measure them in a patient sample, because they are sufficiently characterized by, *e.g.*, mass and by affinity characteristics. It is noted that molecular weight and binding properties are characteristic properties of these markers and not limitations on means of detection or isolation. Furthermore, using the methods described herein or other methods known in the art, the absolute identity of the markers can be determined.

Accordingly, in one aspect the invention provides methods for aiding a breast cancer diagnosis, the method comprising: (a) detecting at least one protein marker in a sample, wherein the protein marker is selected from Marker Br 1: 4170 ± 8 Da; Marker Br 2: 21080 ± 42 Da; Marker Br 3: 9339 ± 19 Da; Marker Br 4: 28308 ± 57 Da; Marker Br 5: 28344 ± 57 Da; Marker Br 6: 4148 ± 9 Da; Marker Br 7: 5634 ± 11 Da; Marker Br 8: 6520 ± 13 Da; Marker Br 9: 6955 ± 14 Da; Marker Br 10: 7507 ± 15 Da; Marker Br 11: 9116 ± 16 Da; Marker Br 12: 9453 ± 19 Da; Marker Br 13: 17310 ± 35 Da; Marker Br 14: 89805 ± 449 Da; Marker Br 15: 4256 ± 9 Da; Marker Br 16: 4357 ± 9 Da; Marker Br 17: 4470 ± 9 Da; Marker Br 18: 9292 ± 19 Da; Marker Br 19: 9335 ± 19 Da; Marker Br 20: 11758 ± 24 Da; Marker Br 21: 11776 ± 24 Da; Marker Br 22: 13913 ± 28 Da; Marker Br 23: 17291 ± 35 Da; Marker Br 24: 17419 ± 35 Da; Marker Br 25: 21103 ± 42 Da; Marker Br 26: 40297 ± 81 Da; Marker Br 27: 4488 ± 9 Da; Marker Br 28: 4647 ± 9 Da; Marker Br 29: 80428 ± 402 Da; Marker Br 30: 11757 ± 24 Da; Marker

Br 31: 4487 ± 9 Da; Marker Br 32: 5360 ± 11 Da; Marker Br 33: 11773 ± 24 Da; Marker Br 34: 13977 ± 28 Da; Marker Br 35: 14004 ± 28 Da; Marker Br 36: 51700 ± 259 Da; and Marker Br 37: 153894 ± 769 Da; and (b) correlating the detection of the marker or markers with a probable diagnosis of breast cancer.

In one embodiment, the correlation takes into account the amount of the marker or markers in the sample and/or the frequency of detection of the same marker or markers in a control.

In another embodiment, gas phase ion spectrometry is used for detecting the marker or markers. For example, laser desorption/ionization mass spectrometry can be used.

In another embodiment, laser desorption/ionization mass spectrometry used to detect markers comprises: (a) providing a substrate comprising an adsorbent attached thereto; (b) contacting the sample with the adsorbent; and (c) desorbing and ionizing the marker or markers from the substrate and detecting the desorbed/ionized marker or markers with the mass spectrometer. Any suitable adsorbents can be used to bind one or more markers. For example, the adsorbent on the substrate can be a hydrophilic adsorbent (*e.g.*, silicon oxide), a metal chelating adsorbent, a lectin adsorbent, a cationic adsorbent, an anionic adsorbent.

In another embodiment, an immunoassay can be used for detecting the marker or markers.

In another embodiment, methods further comprise (a) generating data on the sample with the mass spectrometer indicating intensity of signal for mass/charge ratios; (b) transforming the data into computer-readable form; and (c) operating a computer to execute an algorithm, wherein the algorithm determines closeness-of-fit between the computer-readable data and data indicating a diagnosis of breast cancer or a negative diagnosis.

Another embodiment of this invention is the diagnosis or monitoring of disease progress of a patient having breast cancer where any of the above identified markers are determined from a blood serum sample and monitored under mass spectrometry, preferably using a SELDI chip.

In another aspect, the invention provides methods for detecting at least one protein marker in a sample, wherein the marker is selected from: Marker Br 1: 4170 ± 8 Da; Marker Br 2: 21080 ± 42 Da; Marker Br 3: 9339 ± 19 Da; Marker Br 4: 28308 ± 57 Da; Marker Br 5: 28344 ± 57 Da; Marker Br 6: 4148 ± 9 Da; Marker Br 7: 5634 ± 11 Da;

Marker Br 8: 6520 ± 13 Da; Marker Br 9: 6955 ± 14 Da; Marker Br 10: 7507 ± 15 Da; Marker Br 11: 9116 ± 16 Da; Marker Br 12: 9453 ± 19 Da; Marker Br 13: 17310 ± 35 Da; Marker Br 14: 89805 ± 449 Da; Marker Br 15: 4256 ± 9 Da; Marker Br 16: 4357 ± 9 Da; Marker Br 17: 4470 ± 9 Da; Marker Br 18: 9292 ± 19 Da; Marker Br 19: 9335 ± 19 Da; Marker Br 20: 11758 ± 24 Da; Marker Br 21: 11776 ± 24 Da; Marker Br 22: 13913 ± 28 Da; Marker Br 23: 17291 ± 35 Da; Marker Br 24: 17419 ± 35 Da; Marker Br 25: 21103 ± 42 Da; Marker Br 26: 40297 ± 81 Da; Marker Br 27: 4488 ± 9 Da; Marker Br 28: 4647 ± 9 Da; Marker Br 29: 80428 ± 402 Da; Marker Br 30: 11757 ± 24 Da; Marker Br 31: 4487 ± 9 Da; Marker Br 32: 5360 ± 11 Da; Marker Br 33: 11773 ± 24 Da; Marker Br 34: 13977 ± 28 Da; Marker Br 35: 14004 ± 28 Da; Marker Br 36: 51700 ± 259 Da; and Marker Br 37: 153894 ± 769 Da; wherein the method comprises detecting the marker or markers by gas phase ion spectrometry.

In one embodiment, the methods comprise detecting the marker or markers by laser desorption/ionization mass spectrometry.

In another embodiment, the methods further comprise comparing the amount of the detected marker or markers to a control.

In another embodiment, the methods comprise (a) generating data on the sample with the mass spectrometer indicating intensity of signal for mass/charge ratio; (b) transforming the data into computer-readable form; and (c) operating a computer and executing an algorithm that detects signal in the computer-readable data representing the marker or markers.

In another embodiment, laser desorption/ionization mass spectrometry used to detect a marker or markers comprises (a) providing a substrate comprising an adsorbent attached thereto; (b) contacting the sample with the adsorbent; and (c) desorbing and ionizing the marker or markers from the substrate and detecting the desorbed/ionized marker or markers with the mass spectrometer.

In another embodiment, the methods further comprise sample preparation methods which can improve detection resolution of the markers. For example, the sample preparation includes fractionating a sample by size exclusion chromatography and by anion exchange chromatography, and collecting a sample fraction that includes the marker or markers, and performing gas phase ion spectrometry. In another example, the sample preparation includes removing serum albumin from a sample and collecting a sample fraction that includes the marker or markers, contacting a sample with a series of adsorbents (*e.g.* cationic, anionic, metal chelating, *etc.*) in tandem to capture markers and

performing gas phase ion spectrometry at each adsorbent to detect a marker or markers. In another example, the sample preparation includes fractionating a sample by anion exchange chromatography and collecting a sample fraction that includes the marker or markers, contacting a sample with a series of adsorbents (*e.g.* cationic, anionic, metal chelating, *etc.*) in tandem to capture markers and performing gas phase ion spectrometry at each adsorbent to detect a marker or markers. In another example, the sample preparation includes removing the serum albumin from a sample, fractionating the sample by single stranded DNA chromatography, contacting a sample with a series of adsorbents (*e.g.*, cationic, anionic, metal chelating, *etc.*) in tandem to capture markers and performing gas phase ion spectrometry at each adsorbent to detect marker or markers. In another example, the sample preparation includes fractionating the sample by anion exchange chromatography followed by fractionating the anion exchange chromatography fractions by heparin chromatography, contacting the sample fractions with a series of adsorbents (*e.g.*, cationic, anionic, metal chelating, *etc.*) in tandem to capture markers and performing gas phase ion spectrometry at each adsorbent to detect marker or markers. In another example, the sample preparation includes removing the serum albumin from the sample, fractionating the sample by lectin chromatography, contacting a sample with a series of adsorbents (*e.g.*, cationic, anionic, metal chelating, *etc.*) in tandem to capture markers and performing gas phase ion spectrometry at each adsorbent to detect marker or markers. In another example, the sample preparation includes separating biomolecules in a sample by a gel electrophoresis or high performance liquid chromatography ("HPLC") and obtaining a fraction suspected of comprising the marker or markers, contacting a sample with a series of adsorbents (*e.g.* cationic, anionic, metal chelating, *etc.*) in tandem to capture markers and performing gas phase ion spectrometry at each adsorbent to detect a marker or markers.

In another aspect, the invention provides methods for detecting at least one protein marker in a sample, wherein the marker is selected from: Marker Br 1: 4170 ± 8 Da; Marker Br 2: 21080 ± 42 Da; Marker Br 3: 9339 ± 19 Da; Marker Br 4: 28308 ± 57 Da; Marker Br 5: 28344 ± 57 Da; Marker Br 6: 4148 ± 9 Da; Marker Br 7: 5634 ± 11 Da; Marker Br 8: 6520 ± 13 Da; Marker Br 9: 6955 ± 14 Da; Marker Br 10: 7507 ± 15 Da; Marker Br 11: 9116 ± 16 Da; Marker Br 12: 9453 ± 19 Da; Marker Br 13: 17310 ± 35 Da; Marker Br 14: 89805 ± 449 Da; Marker Br 15: 4256 ± 9 Da; Marker Br 16: 4357 ± 9 Da; Marker Br 17: 4470 ± 9 Da; Marker Br 18: 9292 ± 19 Da; Marker Br 19: 9335 ± 19 Da; Marker Br 20: 11758 ± 24 Da; Marker Br 21: 11776 ± 24 Da; Marker Br 22: 13913 ±

28 Da; Marker Br 23: 17291 ± 35 Da; Marker Br 24: 17419 ± 35 Da; Marker Br 25: 21103 ± 42 Da; Marker Br 26: 40297 ± 81 Da; Marker Br 27: 4488 ± 9 Da; Marker Br 28: 4647 ± 9 Da; Marker Br 29: 80428 ± 402 Da; Marker Br 30: 11757 ± 24 Da; Marker Br 31: 4487 ± 9 Da; Marker Br 32: 5360 ± 11 Da; Marker Br 33: 11773 ± 24 Da; Marker Br 34: 13977 ± 28 Da; Marker Br 35: 14004 ± 28 Da; Marker Br 36: 51700 ± 259 Da; and Marker Br 37: 153894 ± 769 Da; wherein the method comprises detecting the marker or markers by an immunoassay.

In another aspect, the invention provides purified proteins selected from: Marker Br 1: 4170 ± 8 Da; Marker Br 2: 21080 ± 42 Da; Marker Br 3: 9339 ± 19 Da; Marker Br 4: 28308 ± 57 Da; Marker Br 5: 28344 ± 57 Da; Marker Br 6: 4148 ± 9 Da; Marker Br 7: 5634 ± 11 Da; Marker Br 8: 6520 ± 13 Da; Marker Br 9: 6955 ± 14 Da; Marker Br 10: 7507 ± 15 Da; Marker Br 11: 9116 ± 16 Da; Marker Br 12: 9453 ± 19 Da; Marker Br 13: 17310 ± 35 Da; Marker Br 14: 89805 ± 449 Da; Marker Br 15: 4256 ± 9 Da; Marker Br 16: 4357 ± 9 Da; Marker Br 17: 4470 ± 9 Da; Marker Br 18: 9292 ± 19 Da; Marker Br 19: 9335 ± 19 Da; Marker Br 20: 11758 ± 24 Da; Marker Br 21: 11776 ± 24 Da; Marker Br 22: 13913 ± 28 Da; Marker Br 23: 17291 ± 35 Da; Marker Br 24: 17419 ± 35 Da; Marker Br 25: 21103 ± 42 Da; Marker Br 26: 40297 ± 81 Da; Marker Br 27: 4488 ± 9 Da; Marker Br 28: 4647 ± 9 Da; Marker Br 29: 80428 ± 402 Da; Marker Br 30: 11757 ± 24 Da; Marker Br 31: 4487 ± 9 Da; Marker Br 32: 5360 ± 11 Da; Marker Br 33: 11773 ± 24 Da; Marker Br 34: 13977 ± 28 Da; Marker Br 35: 14004 ± 28 Da; Marker Br 36: 51700 ± 259 Da; and Marker Br 37: 153894 ± 769 Da.

In another aspect, the invention provides kits comprising: (a) a substrate comprising an adsorbent attached thereto, wherein the adsorbent is capable of retaining at least one protein marker selected from: Marker Br 1: 4170 ± 8 Da; Marker Br 2: 21080 ± 42 Da; Marker Br 3: 9339 ± 19 Da; Marker Br 4: 28308 ± 57 Da; Marker Br 5: 28344 ± 57 Da; Marker Br 6: 4148 ± 9 Da; Marker Br 7: 5634 ± 11 Da; Marker Br 8: 6520 ± 13 Da; Marker Br 9: 6955 ± 14 Da; Marker Br 10: 7507 ± 15 Da; Marker Br 11: 9116 ± 16 Da; Marker Br 12: 9453 ± 19 Da; Marker Br 13: 17310 ± 35 Da; Marker Br 14: 89805 ± 449 Da; Marker Br 15: 4256 ± 9 Da; Marker Br 16: 4357 ± 9 Da; Marker Br 17: 4470 ± 9 Da; Marker Br 18: 9292 ± 19 Da; Marker Br 19: 9335 ± 19 Da; Marker Br 20: 11758 ± 24 Da; Marker Br 21: 11776 ± 24 Da; Marker Br 22: 13913 ± 28 Da; Marker Br 23: 17291 ± 35 Da; Marker Br 24: 17419 ± 35 Da; Marker Br 25: 21103 ± 42 Da; Marker Br 26: 40297 ± 81 Da; Marker Br 27: 4488 ± 9 Da; Marker Br 28: 4647 ± 9 Da; Marker Br 29: 80428 ± 402 Da; Marker Br 30: 11757 ± 24 Da; Marker Br 31: 4487 ± 9 Da; Marker

Br 32: 5360 ± 11 Da; Marker Br 33: 11773 ± 24 Da; Marker Br 34: 13977 ± 28 Da; Marker Br 35: 14004 ± 28 Da; Marker Br 36: 51700 ± 259 Da; and Marker Br 37: 153894 ± 769 Da; and (b) instructions to detect the marker or markers by contacting a sample with the adsorbent and detecting the marker or markers retained by the adsorbent.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates a sample pre-fractionation protocol, wherein the sample is first pre-fractionated by removing the serum albumin using a Cibacron blue agarose spin column and then further fractionated using single stranded DNA agarose beads in a 96 well filtration plate, and bound to a cationic, anionic, metal-chelating or hydrophobic adsorbent prior to gas phase ion spectrometry analysis.

Figure 2 illustrates a sample pre-fractionation protocol, wherein the sample is pre-fractionated using the Q anion exchange spin column, further fractionated using a heparin spin column, and bound to a cationic, anionic, metal-chelating or hydrophobic adsorbent prior to gas phase ion spectrometry analysis.

Figure 3 illustrates a pre-fractionation protocol, wherein the sample is pre-fractionated by removing serum albumin using an anti-HSA agarose column, further fractionated using a lectin spin column, and bound to a cationic, anionic, metal-chelating or hydrophobic adsorbent prior to gas phase ion spectrometry analysis.

DEFINITIONS

Unless defined otherwise, all technical and scientific terms used herein have the meaning commonly understood by a person skilled in the art to which this invention belongs. The following references provide one of skill with a general definition of many of the terms used in this invention: Singleton *et al.*, *Dictionary of Microbiology and Molecular Biology* (2nd ed. 1994); *The Cambridge Dictionary of Science and Technology* (Walker ed., 1988); *The Glossary of Genetics*, 5th Ed., R. Rieger *et al.* (eds.), Springer Verlag (1991); and Hale & Marham, *The Harper Collins Dictionary of Biology* (1991). As used herein, the following terms have the meanings ascribed to them unless specified otherwise.

"Marker" in the context of the present invention refers to a polypeptide (of a particular apparent molecular weight) which is differentially present in a sample taken from patients having breast cancer as compared to a comparable sample taken from

control subjects (*e.g.*, a person with a negative diagnosis or undetectable cancer, normal or healthy subject).

The phrase "differentially present" refers to differences in the quantity and/or the frequency of a marker present in a sample taken from patients having breast cancer as compared to a control subject. For examples, a marker can be a polypeptide which is present at an elevated level or at a decreased level in samples of breast cancer patients compared to samples of control subjects. Alternatively, a marker can be a polypeptide which is detected at a higher frequency or at a lower frequency in samples of breast cancer patients compared to samples of control subjects. A marker can be differentially present in terms of quantity, frequency or both.

A polypeptide is differentially present between the two samples if the amount of the polypeptide in one sample is statistically significantly different from the amount of the polypeptide in the other sample. For example, a polypeptide is differentially present between the two samples if it is present at least about 120%, at least about 130%, at least about 150%, at least about 180%, at least about 200%, at least about 300%, at least about 500%, at least about 700%, at least about 900%, or at least about 1000% greater than it is present in the other sample, or if it is detectable in one sample and not detectable in the other.

Alternatively or additionally, a polypeptide is differentially present between the two sets of samples if the frequency of detecting the polypeptide in the breast cancer patients' samples is statistically significantly higher or lower than in the control samples. For example, a polypeptide is differentially present between the two sets of samples if it is detected at least about 120%, at least about 130%, at least about 150%, at least about 180%, at least about 200%, at least about 300%, at least about 500%, at least about 700%, at least about 900%, or at least about 1000% more frequently or less frequently observed in one set of samples than the other set of samples.

"Diagnostic" means identifying the presence or nature of a pathologic condition. Diagnostic methods differ in their sensitivity and specificity. The "sensitivity" of a diagnostic assay is the percentage of diseased individuals who test positive (percent of "true positives"). Diseased individuals not detected by the assay are "false negatives." Subjects who are not diseased and who test negative in the assay, are termed "true negatives." The "specificity" of a diagnostic assay is 1 minus the false positive rate, where the "false positive" rate is defined as the proportion of those without the disease who test positive. While a particular diagnostic method may not provide a definitive

diagnosis of a condition, it suffices if the method provides a positive indication that aids in diagnosis.

A "test amount" of a marker refers to an amount of a marker present in a sample being tested. A test amount can be either in absolute amount (*e.g.*, μg/ml) or a relative amount (*e.g.*, relative intensity of signals).

A "diagnostic amount" of a marker refers to an amount of a marker in a subject's sample that is consistent with a diagnosis of breast cancer. A diagnostic amount can be either in absolute amount (*e.g.*, μg/ml) or a relative amount (*e.g.*, relative intensity of signals).

A "control amount" of a marker can be any amount or a range of amount which is to be compared against a test amount of a marker. For example, a control amount of a marker can be the amount of a marker in a person without breast cancer. A control amount can be either in absolute amount (*e.g.*, μg/ml) or a relative amount (*e.g.*, relative intensity of signals).

"Probe" refers to a device that is removably insertable into a gas phase ion spectrometer and comprises a substrate having a surface for presenting a marker for detection. A probe can comprise a single substrate or a plurality of substrates. Terms such as ProteinChip®, ProteinChip® array, or chip are also used herein to refer to specific kinds of probes.

"Substrate" or "probe substrate" refers to a solid phase onto which an adsorbent can be provided (*e.g.*, by attachment, deposition, *etc.*).

"Adsorbent" refers to any material capable of adsorbing a marker. The term "adsorbent" is used herein to refer both to a single material ("monoplex adsorbent") (*e.g.*, a compound or functional group) to which the marker is exposed, and to a plurality of different materials ("multiplex adsorbent") to which the marker is exposed. The adsorbent materials in a multiplex adsorbent are referred to as "adsorbent species." For example, an addressable location on a probe substrate can comprise a multiplex adsorbent characterized by many different adsorbent species (*e.g.*, anion exchange materials, metal chelators, or antibodies), having different binding characteristics. Substrate material itself can also contribute to adsorbing a marker and may be considered part of an "adsorbent."

"Adsorption" or "retention" refers to the detectable binding between an absorbent and a marker either before or after washing with an eluant (selectivity threshold modifier) or a washing solution.

"Eluant" or "washing solution" refers to an agent that can be used to mediate adsorption of a marker to an adsorbent. Eluants and washing solutions are also referred to as "selectivity threshold modifiers." Eluants and washing solutions can be used to wash and remove unbound materials from the probe substrate surface.

"Resolve," "resolution," or "resolution of marker" refers to the detection of at least one marker in a sample. Resolution includes the detection of a plurality of markers in a sample by separation and subsequent differential detection. Resolution does not require the complete separation of one or more markers from all other biomolecules in a mixture. Rather, any separation that allows the distinction between at least one marker and other biomolecules suffices.

"Gas phase ion spectrometer" refers to an apparatus that measures a parameter which can be translated into mass-to-charge ratios of ions formed when a sample is volatilized and ionized. Generally ions of interest bear a single charge, and mass-to-charge ratios are often simply referred to as mass. Gas phase ion spectrometers include, for example, mass spectrometers, ion mobility spectrometers, and total ion current measuring devices.

"Mass spectrometer" refers to a gas phase ion spectrometer that includes an inlet system, an ionization source, an ion optic assembly, a mass analyzer, and a detector.

"Laser desorption mass spectrometer" refers to a mass spectrometer which uses laser as means to desorb, volatilize, and ionize an analyte.

"Detect" refers to identifying the presence, absence or amount of the object to be detected.

The terms "polypeptide," "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an analog or mimetic of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers. Polypeptides can be modified, *e.g.*, by the addition of carbohydrate residues to form glycoproteins. The terms "polypeptide," "peptide" and "protein" include glycoproteins, as well as non-glycoproteins.

"Detectable moiety" or a "label" refers to a composition detectable by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, useful labels include ^{32}P, ^{35}S, fluorescent dyes, electron-dense reagents, enzymes (*e.g.*, as commonly used in an ELISA), biotin-streptavidin, dioxigenin, haptens

and proteins for which antisera or monoclonal antibodies are available, or nucleic acid molecules with a sequence complementary to a target. The detectable moiety often generates a measurable signal, such as a radioactive, chromogenic, or fluorescent signal, that can be used to quantify the amount of bound detectable moiety in a sample. Quantitation of the signal is achieved by, *e.g.*, scintillation counting, densitometry, or flow cytometry.

"Antibody" refers to a polypeptide ligand substantially encoded by an immunoglobulin gene or immunoglobulin genes, or fragments thereof, which specifically binds and recognizes an epitope (*e.g.*, an antigen). The recognized immunoglobulin genes include the kappa and lambda light chain constant region genes, the alpha, gamma, delta, epsilon and mu heavy chain constant region genes, and the myriad immunoglobulin variable region genes. Antibodies exist, *e.g.*, as intact immunoglobulins or as a number of well characterized fragments produced by digestion with various peptidases. This includes, *e.g.*, Fab' and $F(ab)'_2$ fragments. The term "antibody," as used herein, also includes antibody fragments either produced by the modification of whole antibodies or those synthesized de novo using recombinant DNA methodologies. It also includes polyclonal antibodies, monoclonal antibodies, chimeric antibodies, humanized antibodies, or single chain antibodies. "Fc" portion of an antibody refers to that portion of an immunoglobulin heavy chain that comprises one or more heavy chain constant region domains, CH_1, CH_2 and CH_3, but does not include the heavy chain variable region.

"Immunoassay" is an assay that uses an antibody to specifically bind an antigen (*e.g.*, a marker). The immunoassay is characterized by the use of specific binding properties of a particular antibody to isolate, target, and/or quantify the antigen.

The phrase "specifically (or selectively) binds" to an antibody or "specifically (or selectively) immunoreactive with," when referring to a protein or peptide, refers to a binding reaction that is determinative of the presence of the protein in a heterogeneous population of proteins and other biologics. Thus, under designated immunoassay conditions, the specified antibodies bind to a particular protein at least two times the background and do not substantially bind in a significant amount to other proteins present in the sample. Specific binding to an antibody under such conditions may require an antibody that is selected for its specificity for a particular protein. For example, polyclonal antibodies raised to marker Br 1 from specific species such as rat, mouse, or human can be selected to obtain only those polyclonal antibodies that are specifically immunoreactive with marker Br 1 and not with other proteins, except for

polymorphic variants and alleles of marker Br 1. This selection may be achieved by subtracting out antibodies that cross-react with marker Br 1 molecules from other species. A variety of immunoassay formats may be used to select antibodies specifically immunoreactive with a particular protein. For example, solid-phase ELISA immunoassays are routinely used to select antibodies specifically immunoreactive with a protein (*see, e.g.*, Harlow & Lane, *Antibodies, A Laboratory Manual* (1988), for a description of immunoassay formats and conditions that can be used to determine specific immunoreactivity). Typically a specific or selective reaction will be at least twice background signal or noise and more typically more than 10 to 100 times background.

"Energy absorbing molecule" or "EAM" refers to a molecule that absorbs energy from an ionization source in a mass spectrometer thereby aiding desorption of analyte, such as a marker, from a probe surface. Depending on the size and nature of the analyte, the energy absorbing molecule can be optionally used. Energy absorbing molecules used in MALDI are frequently referred to as "matrix." Cinnamic acid derivatives, sinapinic acid ("SPA"), cyano hydroxy cinnamic acid ("CHCA") and dihydroxybenzoic acid are frequently used as energy absorbing molecules in laser desorption of bioorganic molecules.

DETAILED DESCRIPTION OF THE INVENTION

The present invention is based upon, in part, the discovery of protein markers that are differentially present in samples of breast cancer patients and control subjects, and the application of this discovery in methods and kits for aiding a breast cancer diagnosis. Some of these protein markers are found at an elevated level and/or more frequently in samples from breast cancer patients compared to a control (*e.g.*, women in whom breast cancer is undetectable). Accordingly, the amount of one or more markers found in a test sample compared to a control, or the mere detection of one or more markers in the test sample provides useful information regarding probability of whether a subject being tested has breast cancer or not.

The protein markers of the present invention have a number of other uses. For example, the markers can be used to screen for compounds that modulate the expression of the markers *in vitro* or *in vivo*, which compounds in turn may be useful in treating or preventing breast cancer in patients. In another example, markers can be used to monitor responses to certain treatments of breast cancer. In yet another example, the markers can be used in the heredity studies. For instance, certain markers may be

genetically linked. This can be determined by, *e.g.*, analyzing samples from a population of breast cancer patients whose families have a history of breast cancer. The results can then be compared with data obtained from, *e.g.*, breast cancer patients whose families do not have a history of breast cancer. The markers that are genetically linked may be used as a tool to determine if a subject whose family has a history of breast cancer is predisposed to having breast cancer.

I. Characterization of Markers

Three sets of markers were identified from blood serum samples of breast cancer patients using gas phase ion spectrometry. Each set was identified using different pre-fractionation protocols and different adsorbents as described in detail below.

A. Marker Set 1

A first set of markers was found by using a fractionation protocol shown in Figure 1. As shown in Figure 1, a blood serum sample was first diluted into a 8 M urea, 1% CHAPS, 1x PBS buffer and fractionated by first passing the sample over a Cibacron Blue agarose (Sigma, St. Louis, MO) spin column to remove serum albumin, and collecting the flow through. The spin column was washed with 1 M urea, 0.1% CHAPS, 1x PBS and the flow through and wash were combined. 25 µl of albumin depleted sample was added to 100µl of 1M urea, 0.1% CHAPS, 50 mM Tris HCl pH 8.0 buffer and the sample was further fractionated on single stranded DNA agarose (Life Technologies) beads packed in a 96 well filtration plate (Loprodyne membrane, 0.45µm pore size). The single stranded DNA agarose filtration plate was eluted sequentially with the following buffers: buffer 1 containing 1 M urea, 0.1% CHAPS, 50 mM Tris HCl, pH 8, creating fraction D1; buffer 2 containing 50 mM Tris HCl, pH 8, 2.5 M NaCl, 0.1% OGP (n-octyl-β-D-glucopyranoside), creating fraction D2; buffer 3 containing 1.5 M guanidine thiocyanate (GTC), 0.25% OGP, 50 mM HEPES pH 7, creating fraction D3; and buffer 4 containing 33.3% isopropanol, 16.7% acetonitrile, 0.1% trifluroacetic acid (TFA) which was added to fraction D3. Each fraction (D1-D3) was diluted into the proper buffer and applied to a substrate comprising one of the following four adsorbents: a cationic adsorbent (SAX2 ProteinChip® array); a metal adsorbent (IMAC3 Cu(II) ProteinChip® array); a hydrophobic adsorbent (H4 ProteinChip® array); or an anionic adsorbent (WCX2 ProteinChip® array). Each substrate containing markers bound to the indicated adsorbent was tested by gas phase ion spectrometry to measure an apparent

molecular weight of markers retained on the adsorbent. These markers were present at an altered level in breast cancer patients' blood serum samples compared to control samples. Table 1 below shows apparent molecular weights of each marker, fraction in which the markers were found and adsorbents to which the markers bound.

TABLE 1

Marker	App. M.W.	Fraction containing markers and adsorbent binding the marker
Br 1	4170 ± 8 Da	D1 fraction, SAX 2 chip, WCX 2 chip
Br 2	21080 ± 42 Da	D1 fraction, SAX 2 chip
Br 3	9339 ± 19 Da	D1 fraction, IMAC Cu(II) chip
Br 4	28308 ± 57 Da	D1 fraction, IMAC Cu(II) chip
Br 5	28344 ± 57 Da	D1 fraction, IMAC Cu(II) chip
Br 6	4148 ± 9 Da	D1 fraction, WCX 2 chip
Br 7	5634 ± 11 Da	D1 fraction, WCX 2 chip
Br 8	6520 ± 13 Da	D1 fraction, WCX 2 chip
Br 9	6955 ± 14 Da	D1 fraction, WCX 2 chip
Br 10	7507 ± 15 Da	D1 fraction, WCX 2 chip
Br 11	9116 ± 16 Da	D1 fraction, WCX 2 chip
Br 12	9453 ± 19 Da	D1 fraction, WCX 2 chip
Br 13	17310 ± 35 Da	D1 fraction, WCX 2 chip
Br 14	89805 ± 449 Da	D2 fraction, IMAC Cu(II) chip
Br 15	4256 ± 9 Da	D2 fraction, WCX 2 chip
Br 16	4357 ± 9 Da	D2 fraction, WCX 2 chip
Br 17	4470 ± 9 Da	D2 fraction, WCX 2 chip
Br 18	9292 ± 19 Da	D2 fraction, WCX 2 chip
Br 19	9335 ± 19 Da	D2 fraction, WCX 2 chip

As shown in Table 1, the apparent molecular weight of each marker is represented as a range, because the molecular weight of a protein is typically resolved with confidence of about 0.2-0.5% variation by gas phase ion spectrometry. For example, an apparent molecular weight of a protein (having a molecular weight of less than 50,000 daltons) measured by gas phase ion spectrometry can vary about ± 0.2%. For a protein

having a higher molecular weight (*e.g.*, 50,000 daltons or above), an apparent molecular weight of a protein measured by gas phase ion spectrometry can vary about ± 0.5%.

As shown in Table 1, each marker was typically found in only one fraction when a blood serum sample was subject to combination of a Cibacron blue agarose spin column, which removes serum albumin and other proteins (*e.g.*, kinases, dehydrogenases, and other enzymes requiring adenyl-containing cofactors, lipoproteins, blood coagulation factors, interferons, *etc.*), followed by binding to single strand DNA agarose. As shown in Figure 1, fractions "D1" through "D3" refer to three different fractions obtained by eluting sample from the DNA agarose using a series of buffers having different pH, salt and denaturant concentrations. For example, D1 fraction was eluted with a buffer having similar components to the loading buffer. D2 fraction was eluted with a buffer having high ionic strength and a pH of about 8. D3 fraction was eluted with a buffer also having a denaturant and with a buffer containing isopropanol and TFA. Following elution from the single stranded DNA column, fractions D1-D3 were diluted into the proper buffers and bound to an adsorbent surface.

The binding and elution characteristics of the markers provide information regarding markers' physical characteristics. For example, the first set of markers was bound to a single stranded DNA column and was subsequently eluted. Accordingly, the markers likely comprise positively charged or basic moieties. Markers that were eluted with an eluant similar to the loading buffer would likely be loosely bound to the column compared to markers that were eluted with an eluant having high ionic strength or a strong denaturant. These latter eluants will elute markers which are more tightly bound to the column. Moreover, the first set of markers was retained by a cationic adsorbent (SAX2 ProteinChip® array), a transition metal adsorbent (IMAC3 Cu(II) ProteinChip® array), or an anionic adsorbent (WCX2 ProteinChip® array) (*see* the Example section). The cationic adsorbent, (SAX2 ProteinChip® array) comprises a quaternary ammonium surface to bind anionic proteins. The binding occurs through electrostatic interactions of negatively charged amino acids such as aspartic acid, and glutamic acid with the chip surface. Binding occurs at high pH and low salt and binding decreases as pH decreases and salt concentration increases. The transition metal adsorbent (IMAC3 Cu(II) ProteinChip® array) comprises a nitriloacetic acid surface for high capacity copper binding. The binding of proteins occurs through the interaction of proteins with metal binding residues such as exposed histidine, tryptophan and/or cysteine residues with the chip surface. Binding occurs under pH 6-8 and high salt and binding decreases as the

concentration of imidizole and glycine increase. The anionic adsorbent (WCX2 ProteinChip® array) comprises a carboxylate surface to bind cationic proteins. The binding occurs through electrostatic interactions of positively charged amino acids such as lysine, arginine, and histidine with the chip surface. Binding occurs at low pH and low salt, and binding decreases as pH and salt concentration increase. Thirteen of nineteen (68%) markers identified in Marker Set 1 bound to the anionic (WCX2) chip indicating that a majority of the proteins bound to the single stranded DNA column contained positively charged moieties. Four of nineteen (21%) markers identified in Marker Set 1 contained metal binding moieties and two of nineteen (11%) markers identified in Marker Set 1 contained negatively charged moieties.

B. Marker Set 2

A second set of markers was found by using a fractionation protocol shown in Figure 2. As shown in Figure 2, a blood serum sample was first diluted into an 8 M urea, 1% CHAPS, 0.1M Tris buffer pH 9 and fractionated by anion exchange (Q HyperD F, Biosepra, France) chromatography. The anion exchange beads were packed in a 96 well filter plate (Loprodyne membrane, 0.45µm pore size) washed with 100 mM sodium bicarbonate, pH 8.2 and eluted sequentially with the following buffers: buffer 1 containing 100 mM ammonium acetate, pH 7 creating fraction Q1; buffer 2 containing 100 mM sodium acetate, pH 5, creating fraction Q2; buffer 3 containing 100 mM sodium citrate, pH 3, creating fraction Q3; and buffer 4 containing 33.3% isopropanol, 16.7% acetonitrile, 0.1% trifluroacetic acid (TFA) creating fraction Q4. Fractions Q1 and Q2 were combined and fractions Q3 and Q4 were combined. 50 µl of the Q1/Q2 fractions were added to 75 µl of 2 M urea, 0.25% CHAPS, PBS pH 7.2. 40 µl of the Q3/Q4 fractions were added to 75µl 2 M urea, 0.5% CHAPS, 50 mM Tris HCl pH 9 Q1/Q2 and Q3/Q4 fractions were then further fractionated by binding the sample to a heparin bead (Heparin HyperD M, Biosepra, France) 96 well filtration plate (Loprodyne membrane, 0.45µm pore size) and eluting with the following buffers: buffer 1 containing 50 mM Tris HCl, pH 7.2, 0.1% OGP, 0.5M NaCl creating fraction H1; buffer 2 containing 0.1 M Tris HCl, pH 7.2, 2.5 M NaCl, 0.1% OGP, creating fraction H2; buffer 3 containing 1.5 M guanidine thiocyanate (GTC), 0.25% OGP, 50 mM HEPES pH 7 creating fraction H3; and buffer 4 containing 33.3% isopropanol, 16.7% acetonitrile, 0.1% trifluroacetic acid (TFA) and added to fraction H3. Each fraction (H1-H3) was diluted into the proper buffer and applied to a substrate comprising one of the following four adsorbents: a

cationic adsorbent (SAX2 ProteinChip® array); a metal adsorbent (IMAC3 Cu(II) ProteinChip® array); a hydrophobic adsorbent; or an anionic adsorbent (WCX2 ProteinChip® array). Each substrate containing markers bound to the indicated adsorbent was tested by gas phase ion spectrometry to measure an apparent molecular weight of markers retained on the adsorbent. These markers were present at an altered level in breast cancer patients' blood serum samples compared to control samples. Table 2 below shows apparent molecular weights of each marker and fractions in which the markers were found.

TABLE 2

Marker	App. M.W.	Fraction containing markers and adsorbent binding the marker
Br 20	11758 ± 24 Da	Q3 or Q4 fractions, H1 fraction, IMAC Cu(II) chip
Br 21	11776 ± 24 Da	Q3 or Q4 fractions, H1 fraction, IMAC Cu(II) chip
Br 22	13913 ± 28 Da	Q3 or Q4 fractions, H1 fraction, SAX2 chip
Br 23	17291 ± 35 Da	Q3 or Q4 fractions, H1 fraction, SAX2 chip
Br 24	17419 ± 35 Da	Q3 or Q4 fractions, H1 fraction, SAX2 chip
Br 25	21103 ± 42 Da	Q3 or Q4 fractions, H1 fraction, SAX2 chip
Br 26	40297 ± 81 Da	Q3 or Q4 fractions, H1 fraction, SAX2 chip
Br 27	4488 ± 9 Da	Q3 or Q4 fractions, H1 fraction, SAX2 chip
Br 28	4647 ± 9 Da	Q3 or Q4 fractions, H1 fraction, SAX2 chip
Br 29	80428 ± 402 Da	Q3 or Q4 fractions, H1 fraction, SAX2 chip
Br 30	11757± 24 Da	Q3 or Q4 fractions, H2 fraction, IMAC Cu(II) chip
Br 31	4487± 9 Da	Q3 or Q4 fractions, H2 fraction, IMAC Cu(II) chip
Br 32	5359± 11 Da	Q3 or Q4 fractions, H2 fraction, IMAC Cu(II) chip
Br 33	11773 ± 24 Da	Q3 or Q4 fractions, H2 fraction, IMAC Cu(II) chip

As shown in Table 2, each marker in Marker Set 2 can be obtained using the indicated pre-fractionation protocol. The fractionation protocol uses an anion exchange column to bind markers comprising negatively charged moieties. Markers eluted from the column by a buffer with a low pH (*e.g.* fraction Q3) would likely be strongly negatively charged compared to markers eluted with a buffer having a higher pH (*e.g.* Q1 or Q2 fractions). The fractionation protocol also uses a heparin spin column to further fractionate markers with positively charged moieties. Markers eluted from the

heparin column with a buffer having low pH or low salt would likely be weakly positively charged and markers eluted from the heparin column with high salt would likely be strongly positively charged. The markers in Marker Set 2 were retained by a cationic adsorbent (SAX2 ProteinChip® array), a transition metal adsorbent (IMAC3 Cu(II) ProteinChip® array), or an anionic adsorbent (WCX2 ProteinChip® array). Eight of fourteen (57%) markers identified in Marker Set 2 bound to the cationic (SAX2) chip with the remainder of the markers (43%) binding to the metal adsorbent (IMAC3 Cu(II)) chip. A majority of the proteins in Marker Set 2 initially selected for binding to an anion exchange column likely possess negatively charged moieties with the remainder possessing metal binding moieties.

C. **Marker Set 3**

A third set of markers was found by using a fractionation protocol shown in Figure 3. As shown in Figure 3, a blood serum sample was first diluted into a 8M urea/1% CHAPS/1x PBS buffer and fractionated by first passing the sample over an anti-HSA agarose (human serum albumin) spin column, and collecting the flow through. The spin column was washed with 1M urea/0.1% CHAPS/1X PBS and the flow through and wash were combined. 50 µl of the albumin depleted sample was combined with 100 µl of 0.1 M Tris HCl, 0.1% OGP, 0.5 M NaCl pH 7.2 and the sample was further fractionated on lectin beads (wheat germ agglutinin agarose, Sigma, St. Louis, MO) in a 96 well filtration plate (Loprodyne membrane, 0.45µm pore size). The flow through of the lectin beads is fraction W1. The lectin beads were eluted sequentially with the following buffers: buffer 1 containing 0.1 M Tris HCl pH 7.2, 0.1% OGP, 0.5 M NaCl pooled with W1; buffer 2 containing 25 mM N-acetyl-glucosamine/0.1% OGP/1X PBS creating fraction W2; buffer 3 containing 100 mM sodium borate pH 6.5, 0.1% OGP creating fraction W3; and buffer 4 containing 100 mM citrate phosphate pH 3, 0.1% OGP combined into fraction W3. Each fraction (W1-W3) was diluted into the proper buffer and applied to a substrate comprising one of the following four adsorbents: a cationic adsorbent (SAX2 ProteinChip® array); a metal adsorbent (IMAC3 Cu(II) ProteinChip® array); a hydrophobic adsorbent; or an anionic adsorbent (WCX 2 ProteinChip® array). Each substrate containing markers bound to the indicated adsorbent was tested by gas phase ion spectrometry to measure an apparent molecular weight of markers retained on the adsorbent. These markers were present at an altered level in breast cancer patients' blood serum samples compared to control samples. Table 3 below shows apparent

molecular weights of each marker, fraction in which the markers were found and adsorbents to which the markers bound.

TABLE 3

Marker	App. M.W.	Fraction containing markers and adsorbent binding the marker
Br 34	13977 ± 28 Da	W1 fraction, SAX 2 chip
Br 35	14004 ± 28 Da	W1 fraction, SAX 2 chip
Br 36	51700 ± 259 Da	W2 fraction, IMAC Cu(II) chip
Br 37	153894 ± 769 Da	W2 fraction, IMAC Cu(II) chip

As shown in Table 3, each marker in Marker Set 3 can be obtained using the pre-fractionation protocol indicated. The fractionation protocol uses an anti-HSA agarose spin column to remove serum albumin, followed by binding to lectin beads to bind markers which are glycosylated. As shown in Figure 3, fractions "W1" through "W3" refer to three different fractions obtained by eluting sample from the lectin column using a series of buffers having the ability to differentiate the sugar content of the markers. For example, W1 fraction was eluted with a buffer containing salt at a neutral pH. This buffer would remove markers from the lectin column which did not contain any sugar residues. W2 fraction was eluted with a buffer containing a specific sugar that binds to wheat germ lectin. This buffer would remove sugar containing markers from the lectin column. W3 fraction was eluted at low pH and a borate containing buffer which competes with lectin for sugar binding. Following elution from the lectin column, fractions W1-W3 were diluted into the proper buffers and bound to an adsorbent surface as indicated in Figure 3. Two of the four markers identified in Marker Set 3 were bound to the SAX2 chip and two of the four markers were bound to the IMAC3 Cu(II) chip.

While the markers were initially identified from a blood serum sample, the markers may be present in other types of samples (*e.g.*, nipple aspirate, urine, saliva, *etc.*). Thus, samples from which the markers can be detected are not limited to a blood serum sample. Moreover, while the markers were initially identified using the techniques described above, the detection of the markers are not limited by these techniques and other techniques (*e.g.*, immunoassays) can be used.

II. DETECTION OF MARKERS

In another aspect, the invention provides methods for detecting markers which are differentially present in the samples of a breast cancer patient and a control (*e.g.*, women in whom breast cancer is undetectable). The markers can be detected in a number of biological samples. The sample is preferably a biological fluid sample. Examples of a biological fluid sample useful in this invention include blood, blood serum, plasma, nipple aspirate, urine, tears, saliva, *etc.* Because all of the markers are found in blood serum, blood serum is a preferred sample source for embodiments of the invention.

Any suitable methods can be used to detect one or more of the markers described herein. These methods include, without limitation, mass spectrometry (*e.g.*, laser desorption/ionization mass spectrometry), fluorescence (*e.g.* sandwich immunoassay), surface plasmon resonance, ellipsometry and atomic force microscopy.

A. Preparation of a Sample Prior to Detection of Biomarkers

Preferably, the sample is prepared prior to detection of biomarkers. Typically, preparation involves fractionation of the sample and collection of fractions determined to contain the biomarkers. Methods of pre-fractionation include, for example, size exclusion chromatography, ion exchange chromatography, heparin chromatography, affinity chromatography, sequential extraction, gel electrophoresis and liquid chromatography. The analytes also may be modified prior to detection. These methods are useful to simplify the sample for further analysis. For example, it can be useful to remove high abundance proteins, such as albumin, from blood before analysis.

1. Size Exclusion Chromatography

In one embodiment, a sample can be pre-fractionated according to size of proteins in a sample using size exclusion chromatography. For a biological sample wherein the amount of sample available is small, preferably a size selection spin column is used. For example, a K30 spin column (available from Princeton Separation, Ciphergen Biosystems, Inc., *etc.*) can be used. In general, the first fraction that is eluted from the column ("fraction 1") has the highest percentage of high molecular weight proteins; fraction 2 has a lower percentage of high molecular weight proteins; fraction 3 has even a lower percentage of high molecular weight proteins; fraction 4 has the lowest amount of large proteins; and so on. Each fraction can then be analyzed by gas phase ion spectrometry for the detection of markers.

2. Anion Exchange Chromatography

In another embodiment, a sample can be pre-fractionated by anion exchange chromatography. Anion exchange chromatography allows pre-fractionation of the proteins in a sample roughly according to their charge characteristics. For example, a Q anion-exchange resin can be used (*e.g.*, Q HyperD F, Biosepra), and a sample can be sequentially eluted with eluants having different pH's (*see* Figure 2 and Example section VI B). Anion exchange chromatography allows separation of biomolecules in a sample that are more negatively charged from other types of biomolecules. Proteins that are eluted with an eluant having a high pH is likely to be weakly negatively charged, and a fraction that is eluted with an eluant having a low pH is likely to be strongly negatively charged. Thus, in addition to reducing complexity of a sample, anion exchange chromatography separates proteins according to their binding characteristics.

3. Heparin Chromatography

In yet another embodiment, a sample can be pre-fractionated by heparin chromatography. Heparin chromatography allows pre-fractionation of the markers in a sample also on the basis of affinity interaction with heparin and charge characteristics. Heparin, a sulfated mucopolysaccharide, will bind markers with positively charged moieties and a sample can be sequentially eluted with eluants having different pH's or salt concentrations. Markers eluted with an eluant having a low pH are more likely to be weakly positively charged. Markers eluted with an eluant having a high pH are more likely to be strongly positively charged. Thus, heparin chromatography also reduces the complexity of a sample and separates markers according to their binding characteristics.

4. Affinity Chromatography

In yet another embodiment, a sample can be pre-fractionated by removing proteins that are present in a high quantity or that may interfere with the detection of markers in a sample. For example, in a blood serum sample, serum albumin is present in a high quantity and may obscure the analysis of markers. Thus, a blood serum sample can be pre-fractionated by removing serum albumin. Serum albumin can be removed using a substrate that comprises adsorbents that specifically bind serum albumin. For example, a column which comprises, *e.g.*, Cibacron blue agarose (which has a high affinity for serum albumin) or anti-serum albumin antibodies can be used (*see*, *e.g.*, Figures 1 and 3).

In yet another embodiment, a sample can be pre-fractionated by isolating proteins that have a specific characteristic, *e.g.* are glycosylated. For example, a blood serum sample can be fractionated by passing the sample over a lectin chromatography column (which has a high affinity for sugars). Glycosylated proteins will bind to the lectin column and non-glycosylated proteins will pass through the flow through. Glycosylated proteins are then eluted from the lectin column with an eluant containing a sugar, *e.g.*, N-acetyl-glucosamine and are available for further analysis.

Many types of affinity adsorbents exist which are suitable for pre-fractionating blood serum samples. An example of one other type of affinity chromatography available to pre-fractionate a sample is a single stranded DNA spin column. These columns bind proteins which are basic or positively charged. Bound proteins are then eluted from the column using eluants containing denaturants or high pH.

Thus there are many ways to reduce the complexity of a sample based on the binding properties of the proteins in the sample, or the characteristics of the proteins in the sample.

5. Sequential Extraction

In yet another embodiment, a sample can be fractionated using a sequential extraction protocol. In sequential extraction, a sample is exposed to a series of adsorbents to extract different types of biomolecules from a sample. For example, a sample is applied to a first adsorbent to extract certain proteins, and an eluant containing non-adsorbent proteins (*i.e.*, proteins that did not bind to the first adsorbent) is collected. Then, the fraction is exposed to a second adsorbent. This further extracts various proteins from the fraction. This second fraction is then exposed to a third adsorbent, and so on.

Any suitable materials and methods can be used to perform sequential extraction of a sample. For example, a series of spin columns comprising different adsorbents can be used. In another example, a multi-well comprising different adsorbents at its bottom can be used. In another example, sequential extraction can be performed on a probe adapted for use in a gas phase ion spectrometer, wherein the probe surface comprises adsorbents for binding biomolecules. In this embodiment, the sample is applied to a first adsorbent on the probe, which is subsequently washed with an eluant. Markers that do not bind to the first adsorbent is removed with an eluant. The markers that are in the fraction can be applied to a second adsorbent on the probe, and so forth. The advantage of performing sequential extraction on a gas phase ion spectrometer probe

is that markers that bind to various adsorbents at every stage of the sequential extraction protocol can be analyzed directly using a gas phase ion spectrometer.

6. Separation of Biomolecules by Gel Electrophoresis

In yet another embodiment, biomolecules in a sample can be separated by high-resolution electrophoresis, *e.g.*, one or two-dimensional gel electrophoresis. A fraction containing a marker can be isolated and further analyzed by gas phase ion spectrometry. Preferably, two-dimensional gel electrophoresis is used to generate two-dimensional array of spots of biomolecules, including one or more markers. *See, e.g.*, Jungblut and Thiede, *Mass Spectr. Rev.* 16:145-162 (1997).

The two-dimensional gel electrophoresis can be performed using methods known in the art. *See, e.g.*, Deutscher ed., *Methods In Enzymology* vol. 182. Typically, biomolecules in a sample are separated by, *e.g.*, isoelectric focusing, during which biomolecules in a sample are separated in a pH gradient until they reach a spot where their net charge is zero (*i.e.*, isoelectric point). This first separation step results in one-dimensional array of biomolecules. The biomolecules in one dimensional array is further separated using a technique generally distinct from that used in the first separation step. For example, in the second dimension, biomolecules separated by isoelectric focusing are further separated using a polyacrylamide gel, such as polyacrylamide gel electrophoresis in the presence of sodium dodecyl sulfate (SDS-PAGE). SDS-PAGE gel allows further separation based on molecular mass of biomolecules. Typically, two-dimensional gel electrophoresis can separate chemically different biomolecules in the molecular mass range from 1000-200,000 Da within complex mixtures.

Biomolecules in the two-dimensional array can be detected using any suitable methods known in the art. For example, biomolecules in a gel can be labeled or stained (*e.g.*, Coomassie Blue or silver staining). If gel electrophoresis generates spots that correspond to the molecular weight of one or more markers of the invention, the spot can be is further analyzed by gas phase ion spectrometry. For example, spots can be excised from the gel and analyzed by gas phase ion spectrometry. Alternatively, the gel containing biomolecules can be transferred to an inert membrane by applying an electric field. Then a spot on the membrane that approximately corresponds to the molecular weight of a marker can be analyzed by gas phase ion spectrometry. In gas phase ion spectrometry, the spots can be analyzed using any suitable techniques, such as MALDI or SELDI (*e.g.*, using ProteinChip® array) as described in detail below.

Prior to gas phase ion spectrometry analysis, it may be desirable to cleave biomolecules in the spot into smaller fragments using cleaving reagents, such as proteases (*e.g.*, trypsin). The digestion of biomolecules into small fragments provides a mass fingerprint of the biomolecules in the spot, which can be used to determine the identity of markers if desired.

7. High Performance Liquid Chromatography

In yet another embodiment, high performance liquid chromatography (HPLC) can be used to separate a mixture of biomolecules in a sample based on their different physical properties, such as polarity, charge and size. HPLC instruments typically consist of a reservoir of mobile phase, a pump, an injector, a separation column, and a detector. Biomolecules in a sample are separated by injecting an aliquot of the sample onto the column. Different biomolecules in the mixture pass through the column at different rates due to differences in their partitioning behavior between the mobile liquid phase and the stationary phase. A fraction that corresponds to the molecular weight and/or physical properties of one or more markers can be collected. The fraction can then be analyzed by gas phase ion spectrometry to detect markers. For example, the spots can be analyzed using either MALDI or SELDI (*e.g.*, using ProteinChip® array) as described in detail below.

8. Modification of Marker Before Analysis

Optionally, a marker can be modified before analysis to improve its resolution or to determine its identity. For example, the markers may be subject to proteolytic digestion before analysis. Any protease can be used. Proteases, such as trypsin, that are likely to cleave the markers into a discrete number of fragments are particularly useful. The fragments that result from digestion function as a fingerprint for the markers, thereby enabling their detection indirectly. This is particularly useful where there are markers with similar molecular masses that might be confused for the marker in question. Also, proteolytic fragmentation is useful for high molecular weight markers because smaller markers are more easily resolved by mass spectrometry. In another example, biomolecules can be modified to improve detection resolution. For instance, neuraminidase can be used to remove terminal sialic acid residues from glycoproteins to improve binding to an anionic adsorbent (*e.g.*, cationic exchange ProteinChip® arrays) and to improve detection resolution. In another example, the markers can be modified by the attachment of a tag of particular molecular weight that specifically bind to molecular

markers, further distinguishing them. Optionally, after detecting such modified markers, the identity of the markers can be further determined by matching the physical and chemical characteristics of the modified markers in a protein database (*e.g.*, SwissProt).

B. Capture of Biomarkers on Biochips and other Substrates

After preparation, biomarkers in a sample are typically captured on a substrate for detection. Traditional substrates include antibody-coated 96-well plates or nitrocellulose membranes that are subsequently probed for the presence of proteins. More recently, investigators are making use of protein biochips to capture and detect proteins. Many protein biochips are described in the art. These include, for example, protein biochips produced by Ciphergen Biosystems (Fremont, CA), Packard BioScience Company (Meriden CT), Zyomyx (Hayward, CA) and Phylos (Lexington, MA). In general, protein biochips comprise a substrate having a surface. A capture reagent or adsorbent is attached to the surface of the substrate. Frequently, the surface comprises a plurality of addressable locations, each of which location has the capture reagent bound there. The capture reagent can be a biological molecule, such as a polypeptide or a nucleic acid, which captures other biomolecules in a specific manner. Alternatively, the capture reagent can be a chromatographic material, such as an anion exchange material or a hydrophilic material. Examples of such protein biochips are described in the following patents or patent applications: U.S. patent 6,225,047 (Hutchens and Yip, "Use of retentate chromatography to generate difference maps," May 1, 2001), International publication WO 99/51773 (Kuimelis and Wagner, "Addressable protein arrays," October 14, 1999), International publication WO 00/04389 (Wagner et al., "Arrays of protein-capture agents and methods of use thereof," July 27, 2000), International publication WO 00/56934 (Englert et al., "Continuous porous matrix arrays," September 28, 2000).

Protein biochips produced by Ciphergen Biosystems comprise surfaces having chromatographic or biospecific adsorbents attached thereto at addressable locations. Ciphergen ProteinChip® arrays include NP20, H4, SAX-2, WCX-2, IMAC-3, LSAX-30, LWCX-30, IMAC-40, PS-10 and PS-20. Ciphergen's protein biochips comprise an aluminum substrate in the form of a strip. The surface of the strip is coated with silicon dioxide.

In the case of the NP-20 biochip, silicon oxide functions as a hydrophilic adsorbent to capture hydrophilic proteins.

H4, SAX-2, WCX-2, IMAC-3, PS-10 and PS-20 biochips further comprise a functionalized, cross-linked polymer in the form of a hydrogel physically attached to the surface of the biochip or covalently attached through a silane to the surface of the biochip. The H4 biochip has isopropyl functionalities for hydrophobic binding. The SAX-2 biochip has quarternary ammonium functionalities for anion exchange. The WCX-2 biochip has carboxylate functionalities for cation exchange. The IMAC-3 biochip has copper ions immobilized through nitrilotriacetic acid for coordinate covalent bonding. The PS-10 biochip has carboimidizole functional groups that can react with groups on proteins for covalent binding. The PS-20 biochip has epoxide functional groups for covalent binding with proteins. The PS-series biochips are useful for binding biospecific adsorbents, such as antibodies, receptors, lectins, heparin, Protein A, biotin/streptavadin and the like, to chip surfaces where they function to specifically capture analytes from a sample. The LSAX-30 (anion exchange), LWCX-30 (cation exchange) and IMAC-40 (metal chelate) biochips have functionalized latex beads on their surfaces. Such biochips are further described in: WO 00/66265 (Rich et al. ("Probes for a Gas Phase Ion Spectrometer," November 9, 2000); WO 00/67293 (Beecher et al., "Sample Holder with Hydrophobic Coating for Gas Phase Mass Spectrometer," November 9, 2000). United States patent application 09/908,518, filed July 17, 2001 ("Latex Based Adsorbent Chip," Pohl).

In general, a sample containing the biomarkers is placed on the active surface of a biochip for a sufficient time to allow binding. Then, unbound molecules are washed from the surface using a suitable eluant. In general, the more stringent the eluant, the more tightly the proteins must be bound to be retained after the wash. The retained protein biomarkers now can be detected by appropriate means.

C. Detection of Captured Protein Biomarkers

Analytes captured on the surface of a protein biochip can be detected by any method known in the art. This includes, for example, mass spectrometry, fluorescence, surface plasmon resonance, ellipsometry and atomic force microscopy. Mass spectrometry, and particularly SELDI mass spectrometry, is a particularly useful method for detection of the biomarkers of this invention.

1. Detection of Biomarkers by Mass Spectrometry

Preferably, a laser desorption time-of-flight mass spectrometer is used in embodiments of the invention. In laser desorption mass spectrometry, a substrate or a

probe comprising markers is introduced into an inlet system. The markers are desorbed and ionized into the gas phase by laser from the ionization source. The ions generated are collected by an ion optic assembly, and then in a time-of-flight mass analyzer, ions are accelerated through a short high voltage field and let drift into a high vacuum chamber. At the far end of the high vacuum chamber, the accelerated ions strike a sensitive detector surface at a different time. Since the time-of-flight is a function of the mass of the ions, the elapsed time between ion formation and ion detector impact can be used to identify the presence or absence of markers of specific mass to charge ratio.

a) MALDI

Matrix-assisted laser desorption/ionization mass spectrometry, or MALDI-MS, is a method of mass spectrometry that involves the use of an energy absorbing molecule, frequently called a matrix, for desorbing proteins intact from a probe surface. MALDI is described, for example, in U.S. patent 5,118,937 (Hillenkamp et al.) and U.S. patent 5,045,694 (Beavis and Chait). In MALDI-MS the sample is typically mixed with a matrix material and placed on the surface of an inert probe. Exemplary energy absorbing molecules include cinnamic acid derivatives, sinapinic acid ("SPA"), cyano hydroxy cinnamic acid ("CHCA") and dihydroxybenzoic acid. Other suitable energy absorbing molecules are known to those skilled in this art. The matrix dries, forming crystals that encapsulate the analyte molecules. Then the analyte molecules are detected by laser desorption/ionization mass spectrometry. MALDI-MS is useful for detecting the biomarkers of this invention if the complexity of a sample has been substantially reduced using the preparation methods described above.

b) SELDI

Surface-enhanced laser desorption/ionization mass spectrometry, or SELDI-MS represents an improvement over MALDI for the fractionation and detection of biomolecules, such as proteins, in complex mixtures. SELDI is a method of mass spectrometry in which biomolecules, such as proteins, are captured on the surface of a protein biochip using capture reagents that are bound there. Typically, non-bound molecules are washed from the probe surface before interrogation. SELDI technology is available from Ciphergen Biosystems, Inc., Fremont CA as part of the ProteinChip® System. ProteinChip® arrays are particularly adapted for use in SELDI. SELDI is described, for example, in: United States Patent 5,719,060 ("Method and Apparatus for Desorption and Ionization of Analytes," Hutchens and Yip, February 17, 1998,) United

States Patent 6,225,047 ("Use of Retentate Chromatography to Generate Difference Maps," Hutchens and Yip, May 1, 2001) and Weinberger et al., "Time-of-flight mass spectrometry," in Encyclopedia of Analytical Chemistry, R.A. Meyers, ed., pp 11915-11918 John Wiley & Sons Chichesher, 2000.

2. Analysis of Desorption/Ionization and Detection

Markers on the substrate surface can be desorbed and ionized using gas phase ion spectrometry. Any suitable gas phase ion spectrometers can be used as long as it allows markers on the substrate to be resolved. Preferably, gas phase ion spectrometers allow quantitation of markers.

In one embodiment, a gas phase ion spectrometer is a mass spectrometer. In a typical mass spectrometer, a substrate or a probe comprising markers on its surface is introduced into an inlet system of the mass spectrometer. The markers are then desorbed by a desorption source such as a laser, fast atom bombardment, high energy plasma, electrospray ionization, thermospray ionization, liquid secondary ion MS, field desorption, *etc.* The generated desorbed, volatilized species consist of preformed ions or neutrals which are ionized as a direct consequence of the desorption event. Generated ions are collected by an ion optic assembly, and then a mass analyzer disperses and analyzes the passing ions. The ions exiting the mass analyzer are detected by a detector. The detector then translates information of the detected ions into mass-to-charge ratios. Detection of the presence of markers or other substances will typically involve detection of signal intensity. This, in turn, can reflect the quantity and character of markers bound to the substrate. Any of the components of a mass spectrometer (*e.g.*, a desorption source, a mass analyzer, a detector, *etc.*) can be combined with other suitable components described herein or others known in the art in embodiments of the invention.

Preferably, a laser desorption time-of-flight mass spectrometer is used in embodiments of the invention. In laser desorption mass spectrometry, a substrate or a probe comprising markers is introduced into an inlet system. The markers are desorbed and ionized into the gas phase by laser from the ionization source. The ions generated are collected by an ion optic assembly, and then in a time-of-flight mass analyzer, ions are accelerated through a short high voltage field and let drift into a high vacuum chamber. At the far end of the high vacuum chamber, the accelerated ions strike a sensitive detector surface at a different time. Since the time-of-flight is a function of the mass of the ions,

the elapsed time between ion formation and ion detector impact can be used to identify the presence or absence of markers of specific mass to charge ratio.

In another embodiment, an ion mobility spectrometer can be used to detect markers. The principle of ion mobility spectrometry is based on different mobility of ions. Specifically, ions of a sample produced by ionization move at different rates, due to their difference in, *e.g.*, mass, charge, or shape, through a tube under the influence of an electric field. The ions (typically in the form of a current) are registered at the detector which can then be used to identify a marker or other substances in a sample. One advantage of ion mobility spectrometry is that it can operate at atmospheric pressure.

In yet another embodiment, a total ion current measuring device can be used to detect and characterize markers. This device can be used when the substrate has a only a single type of marker. When a single type of marker is on the substrate, the total current generated from the ionized marker reflects the quantity and other characteristics of the marker. The total ion current produced by the marker can then be compared to a control (*e.g.*, a total ion current of a known compound). The quantity or other characteristics of the marker can then be determined.

3. Detection by Immunoassay

In another embodiment, an immunoassay can be used to detect and analyze markers in a sample. This method comprises: (a) providing an antibody that specifically binds to a marker; (b) contacting a sample with the antibody; and (c) detecting the presence of a complex of the antibody bound to the marker in the sample.

To prepare an antibody that specifically binds to a marker, purified markers or their nucleic acid sequences can be used. Nucleic acid and amino acid sequences for markers can be obtained by further characterization of these markers. For example, each marker can be peptide mapped with a number of enzymes (*e.g.*, trypsin, V8 protease, *etc.*). The molecular weights of digestion fragments from each marker can be used to search the databases, such as SwissProt database, for sequences that will match the molecular weights of digestion fragments generated by various enzymes. Using this method, the nucleic acid and amino acid sequences of other markers can be identified if these markers are known proteins in the databases.

Alternatively, the proteins can be sequenced using protein ladder sequencing. Protein ladders can be generated by, for example, fragmenting the molecules and subjecting fragments to enzymatic digestion or other methods that sequentially

remove a single amino acid from the end of the fragment. Methods of preparing protein ladders are described, for example, in International Publication WO 93/24834 (Chait *et al.*) and United States Patent 5,792,664 (Chait *et al.*). The ladder is then analyzed by mass spectrometry. The difference in the masses of the ladder fragments identify the amino acid removed from the end of the molecule.

If the markers are not known proteins in the databases, nucleic acid and amino acid sequences can be determined with knowledge of even a portion of the amino acid sequence of the marker. For example, degenerate probes can be made based on the N-terminal amino acid sequence of the marker. These probes can then be used to screen a genomic or cDNA library created from a sample from which a marker was initially detected. The positive clones can be identified, amplified, and their recombinant DNA sequences can be subcloned using techniques which are well known. *See, e.g., Current Protocols for Molecular Biology* (Ausubel *et al.*, Green Publishing Assoc. and Wiley-Interscience 1989) *and Molecular Cloning: A Laboratory Manual*, 3rd Ed. (Sambrook *et al.*, Cold Spring Harbor Laboratory, NY 2001).

Using the purified markers or their nucleic acid sequences, antibodies that specifically bind to a marker can be prepared using any suitable methods known in the art. *See, e.g.*, Coligan, *Current Protocols in Immunology* (1991); Harlow & Lane, *Antibodies: A Laboratory Manual* (1988); Goding, *Monoclonal Antibodies: Principles and Practice* (2d ed. 1986); and Kohler & Milstein, *Nature* 256:495-497 (1975). Such techniques include, but are not limited to, antibody preparation by selection of antibodies from libraries of recombinant antibodies in phage or similar vectors, as well as preparation of polyclonal and monoclonal antibodies by immunizing rabbits or mice (*see, e.g.*, Huse *et al.*, *Science* 246:1275-1281 (1989); Ward *et al.*, *Nature* 341:544-546 (1989)).

After the antibody is provided, a marker can be detected and/or quantified using any of suitable immunological binding assays known in the art (*see, e.g.*, U.S. Patent Nos. 4,366,241; 4,376,110; 4,517,288; and 4,837,168). Useful assays include, for example, an enzyme immune assay (EIA) such as enzyme-linked immunosorbent assay (ELISA), a radioimmune assay (RIA), a Western blot assay, or a slot blot assay. These methods are also described in, *e.g.*, *Methods in Cell Biology: Antibodies in Cell Biology*, volume 37 (Asai, ed. 1993); *Basic and Clinical Immunology* (Stites & Terr, eds., 7th ed. 1991); and Harlow & Lane, *supra*.

Generally, a sample obtained from a subject can be contacted with the antibody that specifically binds the marker. Optionally, the antibody can be fixed to a solid support to facilitate washing and subsequent isolation of the complex, prior to contacting the antibody with a sample. Examples of solid supports include glass or plastic in the form of, *e.g.*, a microtiter plate, a stick, a bead, or a microbead. Antibodies can also be attached to a probe substrate or ProteinChip® array described above. The sample is preferably a biological fluid sample taken from a subject. Examples of biological fluid samples include blood, serum, plasma, nipple aspirate, urine, tears, saliva *etc.* In a preferred embodiment, the biological fluid comprises blood serum. The sample can be diluted with a suitable eluant before contacting the sample to the antibody.

After incubating the sample with antibodies, the mixture is washed and the antibody-marker complex formed can be detected. This can be accomplished by incubating the washed mixture with a detection reagent. This detection reagent may be, *e.g.*, a second antibody which is labeled with a detectable label. Exemplary detectable labels include magnetic beads (*e.g.*, DYNABEADS™), fluorescent dyes, radiolabels, enzymes (*e.g.*, horse radish peroxide, alkaline phosphatase and others commonly used in an ELISA), and colorimetric labels such as colloidal gold or colored glass or plastic beads. Alternatively, the marker in the sample can be detected using an indirect assay, wherein, for example, a second, labeled antibody is used to detect bound marker-specific antibody, and/or in a competition or inhibition assay wherein, for example, a monoclonal antibody which binds to a distinct epitope of the marker is incubated simultaneously with the mixture.

Throughout the assays, incubation and/or washing steps may be required after each combination of reagents. Incubation steps can vary from about 5 seconds to several hours, preferably from about 5 minutes to about 24 hours. However, the incubation time will depend upon the assay format, marker, volume of solution, concentrations and the like. Usually the assays will be carried out at ambient temperature, although they can be conducted over a range of temperatures, such as 10°C to 40°C.

Immunoassays can be used to determine presence or absence of a marker in a sample as well as the quantity of a marker in a sample. First, a test amount of a marker in a sample can be detected using the immunoassay methods described above. If a marker is present in the sample, it will form an antibody-marker complex with an

antibody that specifically binds the marker under suitable incubation conditions described above. The amount of an antibody-marker complex can be determined by comparing to a standard. A standard can be, *e.g.*, a known compound or another protein known to be present in a sample. As noted above, the test amount of marker need not be measured in absolute units, as long as the unit of measurement can be compared to a control.

The methods for detecting these markers in a sample have many applications. For example, one or more markers can be measured to aid breast cancer diagnosis or prognosis. In another example, the methods for detection of the markers can be used to monitor responses in a subject to cancer treatment. In another example, the methods for detecting markers can be used to assay for and to identify compounds that modulate expression of these markers *in vivo* or *in vitro*.

D. Analysis of Data

Data generated by desorption and detection of markers can be analyzed using any suitable means. In one embodiment, data is analyzed with the use of a programmable digital computer. The computer program generally contains a readable medium that stores codes. Certain code can be devoted to memory that includes the location of each feature on a probe, the identity of the adsorbent at that feature and the elution conditions used to wash the adsorbent. The computer also contains code that receives as input, data on the strength of the signal at various molecular masses received from a particular addressable location on the probe. This data can indicate the number of markers detected, including the strength of the signal generated by each marker.

Data analysis can include the steps of determining signal strength (*e.g.*, height of peaks) of a marker detected and removing "outliers" (data deviating from a predetermined statistical distribution). The observed peaks can be normalized, a process whereby the height of each peak relative to some reference is calculated. For example, a reference can be background noise generated by instrument and chemicals (*e.g.*, energy absorbing molecule) which is set as zero in the scale. Then the signal strength detected for each marker or other biomolecules can be displayed in the form of relative intensities in the scale desired (*e.g.*, 100). Alternatively, a standard (*e.g.*, a serum protein) may be admitted with the sample so that a peak from the standard can be used as a reference to calculate relative intensities of the signals observed for each marker or other markers detected.

The computer can transform the resulting data into various formats for displaying. In one format, referred to as "spectrum view or retentate map," a standard spectral view can be displayed, wherein the view depicts the quantity of marker reaching the detector at each particular molecular weight. In another format, referred to as "peak map," only the peak height and mass information are retained from the spectrum view, yielding a cleaner image and enabling markers with nearly identical molecular weights to be more easily seen. In yet another format, referred to as "gel view," each mass from the peak view can be converted into a grayscale image based on the height of each peak, resulting in an appearance similar to bands on electrophoretic gels. In yet another format, referred to as "3-D overlays," several spectra can be overlaid to study subtle changes in relative peak heights. In yet another format, referred to as "difference map view," two or more spectra can be compared, conveniently highlighting unique markers and markers which are up- or down-regulated between samples. Marker profiles (spectra) from any two samples may be compared visually. In yet another format, Spotfire Scatter Plot can be used, wherein markers that are detected are plotted as a dot in a plot, wherein one axis of the plot represents the apparent molecular of the markers detected and another axis represents the signal intensity of markers detected. For each sample, markers that are detected and the amount of markers present in the sample can be saved in a computer readable medium. This data can then be compared to a control (*e.g.*, a profile or quantity of markers detected in control, *e.g.*, women in whom breast cancer is undetectable).

III. DIAGNOSIS OF BREAST CANCER

In another aspect, the invention provides methods for aiding a breast cancer diagnosis using one or more markers in Marker Set 1, Marker Set 2, or Marker Set 3. These markers can be used alone, in combination with other markers in any set, or with entirely different markers (*e.g.*, neu oncogene product) in aiding breast cancer diagnosis. The markers in Marker Set 1, Marker Set 2, and Marker Set 3 are differentially present in samples of a breast cancer patient and a normal subject in whom breast cancer is undetectable. For example, some of the markers are expressed at an elevated level and/or are present at a higher frequency in breast cancer patients than in normal subjects. Therefore, detection of one or more of these markers in a person would provide useful information regarding the probability that the person may have breast cancer.

Accordingly, embodiments of the invention include methods for aiding a breast cancer diagnosis, wherein the method comprises: (a) detecting at least one marker in a sample, wherein the marker is selected from Marker Br 1: 4170 ± 8 Da; Marker Br 2: 21080 ± 42 Da; Marker Br 3: 9339 ± 19 Da; Marker Br 4: 28308 ± 57 Da; Marker Br 5: 28344 ± 57 Da; Marker Br 6: 4148 ± 9 Da; Marker Br 7: 5634 ± 11 Da; Marker Br 8: 6520 ± 13 Da; Marker Br 9: 6955 ± 14 Da; Marker Br 10: 7507 ± 15 Da; Marker Br 11: 9116 ± 16 Da; Marker Br 12: 9453 ± 19 Da; Marker Br 13: 17310 ± 35 Da; Marker Br 14: 89805 ± 449 Da; Marker Br 15: 4256 ± 9 Da; Marker Br 16: 4357 ± 9 Da; Marker Br 17: 4470 ± 9 Da; Marker Br 18: 9292 ± 19 Da; Marker Br 19: 9335 ± 19 Da; Marker Br 20: 11758 ± 24 Da; Marker Br 21: 11776 ± 24 Da; Marker Br 22: 13913 ± 28 Da; Marker Br 23: 17291 ± 35 Da; Marker Br 24: 17419 ± 35 Da; Marker Br 25: 21103 ± 42 Da; Marker Br 26: 40297 ± 81 Da; Marker Br 27: 4488 ± 9 Da; Marker Br 28: 4647 ± 9 Da; Marker Br 29: 80428 ± 402 Da; Marker Br 30: 11757 ± 24 Da; Marker Br 31: 4487 ± 9 Da; Marker Br 32: 5360 ± 11 Da; Marker Br 33: 11773 ± 24 Da; Marker Br 34: 13977 ± 28 Da; Marker Br 35: 14004 ± 28 Da; Marker Br 36: 51700 ± 259 Da; and Marker Br 37: 153894 ± 769 Da; and (b) correlating the detection of the marker or markers with a probable diagnosis of breast cancer. The correlation may take into account the amount of the marker or markers in the sample compared to a control amount of the marker or markers (up or down regulation of the marker or markers) (*e.g.*, in normal subjects in whom breast cancer is undetectable). The correlation may take into account the presence or absence of the markers in a test sample and the frequency of detection of the same markers in a control. The correlation may take into account both of such factors to facilitate determination of whether a subject has a breast cancer or not.

Any suitable samples can be obtained from a subject to detect markers. Preferably, a sample is a blood serum sample from the subject. If desired, the sample can be prepared as described above to enhance detectability of the markers. For example, to increase the detectability of markers Br 1, Br 2, Br 3, Br 4, Br 5, Br 6, Br 7, Br 8, Br 9, Br 10, Br 11, Br 12, Br 13, Br 14, Br 15, Br 16, Br 17, Br 18, and Br 19 a blood serum sample from the subject can be preferably fractionated by, *e.g.*, Cibacron blue agarose chromatography and single stranded DNA affinity chromatography. In another example, if the detection of markers Br 20, Br 21, Br 22, Br 23, Br 24, Br 25, Br 26, Br 27, Br 28, Br 29, Br 30, Br 31, Br 32, and Br 33 is desired, then the sample is preferably pre-fractionated by anion exchange chromatography followed by heparin chromatography. In another example, if the detection of markers Br 34, Br 35, Br 36, and Br 37 is desired, the

sample can be pre-fractionated by removing the serum albumin followed by binding to a lectin column. Sample preparations, such as pre-fractionation protocols, is optional and may not be necessary to enhance detectability of markers depending on the methods of detection used. For example, sample preparation may be unnecessary if antibodies that specifically bind markers are used to detect the presence of markers in a sample.

Any suitable method can be used to detect a marker or markers in a sample. For example, gas phase ion spectrometry or an immunoassay can be used as described above. Using these methods, one or more markers can be detected. Preferably, a sample is tested for the presence of a plurality of markers. Detecting the presence of a plurality of markers, rather than a single marker alone, would provide more information for the diagnostician. Specifically, the detection of a plurality of markers in a sample would increase the percentage of true positive and true negative diagnoses and would decrease the percentage of false positive or false negative diagnoses.

The detection of the marker or markers is then correlated with a probable diagnosis of breast cancer. In some embodiments, the detection of the mere presence or absence of a marker, without quantifying the amount of marker, is useful and can be correlated with a probable diagnosis of breast cancer. For example, markers Br 1, Br 6, Br 7, Br 8, Br 9, Br 10, Br 11, Br 15, Br 16, BR 17, Br 20, Br 27, Br 28, Br 30, Br 31, Br 32, Br 33 and BR 36 are more frequently detected in breast cancer patients than in normal subjects. Thus, a mere detection of one or more of these markers in a subject being tested indicates that the subject has a higher probability of having a breast cancer.

In other embodiments, the detection of markers can involve quantifying the markers to correlate the detection of markers with a probable diagnosis of breast cancer. For example, markers Br 1, Br 6, Br 7, Br 8, Br 9, Br 10, Br 11, Br 15, Br 16, Br 17, Br 20, Br 27, Br 28, Br 30, Br 31, Br 32, Br 33 and BR 36 are present at a higher quantity in blood serum samples of breast cancer patients than in blood serum samples of normal subjects. Thus, if the amount of the markers detected in a subject being tested is higher compared to a control amount, then the subject being tested has a higher probability of having a breast cancer.

Similarly, in another embodiment, the detection of markers can further involve quantifying the markers to correlate the detection of markers with a probable diagnosis of breast cancer wherein the markers are present in lower quantities in blood serum samples from breast cancer patients than in blood serum samples of normal subjects. For example, markers Br 2, Br 3, Br 4, Br 12, Br 13, Br 14, Br 18, Br 19, Br 22,

Br 23, Br 24, Br 25, Br 26, Br 29, Br 35, and BR 37 are present at a lower quantity in blood serum samples of breast cancer patients than in blood serum samples of normal subjects. Thus, if the amount of the markers detected in a subject being tested is lower compared to a control amount, then the subject being tested has a higher probability of having a breast cancer.

An analysis of the data shows that detection of any one of markers Br 1, Br 6, Br 7, Br 8, Br 9, Br 10, Br 11, Br 15, Br 16, Br 17, Br 20, Br 27, Br 28, Br 30, Br 31, Br 32, Br 33 and BR 36 is not highly correlated with a positive diagnosis of breast cancer. However, the chance of a positive diagnosis increases significantly with the detection of any two, three, four or five of these markers. Furthermore, the failure to detect any one, two or any three of these markers also is not highly correlated with a negative diagnosis of breast cancer. However, the failure to detect any two, three, four or five of these markers is highly correlated with a negative diagnosis of breast cancer.

Similarly, an analysis of the data shows that the failure to detect any one of markers Br 2, Br 3, Br 4, Br 12, Br 13, Br 14, Br 18, Br 19, Br 22, Br 23, Br 24, Br 25, Br 26, Br 29, Br 35, and BR 37 is not highly correlated with a positive diagnosis of breast cancer. However, the chance of a positive diagnosis increases significantly by failing to detect any two, three, four or five of these markers. In addition, detecting any one of these markers also is not highly correlated with a negative diagnosis of breast cancer. However, detecting any two, three, four or five of these markers is highly correlated with a negative diagnosis of breast cancer.

When the markers are quantified, it can be compared to a control. A control can be, *e.g.*, the average or median amount of marker present in comparable samples of normal subjects in whom breast cancer is undetectable. The control amount is measured under the same or substantially similar experimental conditions as in measuring the test amount. For example, if a test sample is obtained from a subject's blood serum sample and a marker is detected using a particular probe, then a control amount of the marker is preferably determined from a serum sample of a patient using the same probe. It is preferred that the control amount of marker is determined based upon a significant number of samples from normal subjects who do not have breast cancer so that it reflects variations of the marker amounts in that population.

Data generated by mass spectrometry can then be analyzed by a computer software. The software can comprise code that converts signal from the mass spectrometer into computer readable form. The software also can include code that

applies an algorithm to the analysis of the signal to determine whether the signal represents a "peak" in the signal corresponding to a marker of this invention, or other useful markers. The software also can include code that executes an algorithm that compares signal from a test sample to a typical signal characteristic of "normal" and breast cancer and determines the closeness of fit between the two signals. The software also can include code indicating which the test sample is closest to, thereby providing a probable diagnosis.

IV. KITS

In yet another aspect, the invention provides kits for aiding a diagnosis of breast cancer, wherein the kits can be used to detect the markers of the present invention. For example, the kits can be used to detect any one or more of the markers described herein, which markers are differentially present in samples of a breast cancer patient and normal subjects. The kits of the invention have many applications. For example, the kits can be used to differentiate if a subject has breast cancer or has a negative diagnosis, thus aiding a breast cancer diagnosis. In another example, the kits can be used to identify compounds that modulate expression of one or more of the markers in *in vitro* or *in vivo* animal models for breast cancer.

In one embodiment, a kit comprises: (a) a substrate comprising an adsorbent thereon, wherein the adsorbent is suitable for binding a marker, and (b) instructions to detect the marker or markers by contacting a sample with the adsorbent and detecting the marker or markers retained by the adsorbent. In some embodiments, the kit may comprise an eluant (as an alternative or in combination with instructions) or instructions for making an eluant, wherein the combination of the adsorbent and the eluant allows detection of the markers using gas phase ion spectrometry. Such kits can be prepared from the materials described above, and the previous discussion of these materials (*e.g.*, probe substrates, adsorbents, washing solutions, *etc.*) is fully applicable to this section and will not be repeated.

In another embodiment, the kit may comprise a first substrate comprising an adsorbent thereon (*e.g.*, a particle functionalized with an adsorbent) and a second substrate onto which the first substrate can be positioned to form a probe which is removably insertable into a gas phase ion spectrometer. In other embodiments, the kit may comprise a single substrate which is in the form of a removably insertable probe with adsorbents on the substrate. In yet another embodiment, the kit may further comprise a

pre-fractionation spin column (*e.g.*, Cibacron blue agarose column, anti-HSA agarose column, K-30 size exclusion column, Q-anion exchange spin column, single stranded DNA column, lectin column, *etc.*).

Optionally, the kit can further comprise instructions for suitable operational parameters in the form of a label or a separate insert. For example, the kit may have standard instructions informing a consumer how to wash the probe after a sample of blood serum is contacted on the probe. In another example, the kit may have instructions for pre-fractionating a sample to reduce complexity of proteins in the sample. In another example, the kit may have instructions for automating the fractionation or other processes.

In another embodiment, a kit comprises (a) an antibody that specifically binds to a marker; and (b) a detection reagent. Such kits can be prepared from the materials described above, and the previous discussion regarding the materials (*e.g.*, antibodies, detection reagents, immobilized supports, *etc.*) is fully applicable to this section and will not be repeated. Optionally, the kit may further comprise pre-fractionation spin columns. In some embodiments, the kit may further comprise instructions for suitable operation parameters in the form of a label or a separate insert.

Optionally, the kit may further comprise a standard or control information so that the test sample can be compared with the control information standard to determine if the test amount of a marker detected in a sample is a diagnostic amount consistent with a diagnosis of breast cancer.

EXAMPLES

The following examples are offered by way of illustration, not by way of limitation. Illustrated below are the probe preparation protocols, the sample preparation protocols, and the identification of markers that have positive or negative correlations with breast cancer. While the sample protocols have been developed with human blood serum samples, the same general experimental set-up may be used for suitable samples to detect markers.

V. SERUM SAMPLE PROFILING ON ProteinChip® ARRAYS

A sample can be analyzed using a number of different chips. Preferably, "a multi-well bioprocessor" is used to analyze a number of samples simultaneously. If desired, each sample is washed with two or more different washes per chip type (*e.g.*;

wash 1 and wash 2). These washes vary depending on the surface chemistry of the chip. Each sample can be analyzed using two or more different types of EAM (*e.g.*, sinapinic acid ("SPA") and alpha cyano hydroxy cinnamic acid ("CHCA")). Thus, each sample can be analyzed under various conditions in order to maximize the diversity of proteins, including markers, detected. Samples can also be analyzed in multiple sets to minimize the impact of experimental variability. The protocols described below are merely exemplary and any variation thereof would be readily apparent to one of skill in the art. Accordingly, embodiments of the invention are not limited to the protocols described below.

A. Protein Profiling on SAX2 ProteinChip® Array

All of the following steps can be performed automatically with a robotic station (Biomek 2000Laboratory Automation Workstation, Beckman Coulter, CA).

1. Put SAX2 chips in bioprocessor. Attach the top securely.
2. Add 200 µL 50 mM HEPES, pH 7.4 to each well.
3. Mix, 250 rpm on a programmable reciprocating platform (Micromix5, DPC Cirrus, Inc. NJ), 5 min, RT.
4. Discard buffer.
5. Repeat steps 2-3.
6. To pre-fractionated serum samples: Dilute 10-fold in 50 mM HEPES, pH 7.4.
7. Add 50 µL diluted serum to each well.
8. Mix (250 rpm on a platform shaker) at room temperature for 15 min.
9. Remove the samples from SAX2 chips.
   a. Add 200 µL of 50 mM HEPES pH 7.4 (Wash 1); or
   b. Add 200 µL of 1 M urea 0.125% CHAPS, 0.25 M NaCl, 50 mM HEPES pH 7.4 (Wash 2).
10. Mix (250 rpm on a platform shaker) at room temperature for 5 min.
11. Remove the washes from the spots.
    a. Add fresh 200 µL of 50 mM HEPES (Wash 1); or
    b. Add fresh 200 µL of 1 M urea 0.125% CHAPS, 0.25 M NaCl, 50 mM HEPES (Wash 2).
12. Mix vigorously (250 rpm on a platform shaker) at room temperature for 5 min.

13. Repeat buffer wash one more time.
14. Wash chips 3 times with water by filling the wells and emptying.
15. Remove chips from bioprocessor.
16. Rinse chips with deionized H_2O.
17. Air dry chips.
18. Add 0.5 µL of SPA or CHCA to chips two times (air dry spots between additions).

B. Protein profiling on WCX2 ProteinChip® Array

All of the following steps can be performed automatically with a robotic station (Biomek 2000Laboratory Automation Workstation, Beckman Coulter, CA).

1. Put WCX2 chips in bioprocessor. Attach the top securely.
2. Add 200 µL 50 mM HEPES, pH 7.4 to each well.
3. Mix, 250 rpm on a programmable reciprocating shaking platform (MicroMix5, DPC Cirrus, Inc. NJ), 5 min, RT.
4. Discard buffer.
5. Repeat steps 2-3.
6. To pre-fractionated serum samples: Dilute 10-fold in 50 mM HEPES, pH 7.4.
7. Add 50 µL diluted serum to each well.
8. Mix (250 rpm on a platform shaker) at room temperature for 15 min.
9. Remove the samples from WCX2 chips.
   a. Add 200 µL of 50 mM HEPES pH 7.4 (Wash 1); or
   b. Add 200 µL of 1 M urea 0.125% CHAPS, 0.25 M NaCl, 50 mM HEPES pH 7.4 (Wash 2).
10. Mix (250 rpm on a platform shaker) at room temperature for 5 min.
11. Remove the washes from the spots.
    a. Add fresh 200 µL of 50 mM HEPES (Wash 1); or
    b. Add fresh 200 µL of 1 M urea 0.125% CHAPS, 0.25 M NaCl, 50 mM HEPES (Wash 2).
12. Mix vigorously (250 rpm on a platform shaker) at room temperature for 5 min.
13. Repeat buffer wash one more time.
14. Wash chips 3 times with water by filling the wells and emptying.

15. Remove chips from bioprocessor.
16. Rinse chips with deionized H_2O.
17. Air dry chips.
18. Add 0.5 µL of SPA or CHCA to chips two times (air dry spots between additions).

C. Protein Profiling on IMAC3-Cu^{++} ProteinChip® Array

1. Add 10 µL of 100 mM $CuSO_4$ to each spot on IMAC3 chips.
2. Mix for 5 min, RT.
3. Rinse chips with water.
4. Wash each spot with 10 µL 100 mM sodium acetate pH 4.0 on a shaker for 5 min, RT.
5. Rinse chips with water. All of the following steps can be performed automatically with a robotic station (Biomek 2000Laboratory Automation Workstation, Beckman Coulter, CA)
6. Put chips in Bioprocessor. Attach the top securely.
7. Add 200 µL PBS buffer pH 7.2 to each spot.
8. Mix 250 rpm on a programmable reciprocating shaking platform (MicroMix5, DPC Cirrus, Inc., NJ) at room temperature for 5 min. Discard buffer.
9. Repeat wash with fresh PBS.
10. Dilute the pre-fractionated serum sample 10 fold in PBS.
11. Add 50 µL diluted serum to all spots.
12. Mix vigorously (250 rpm on a platform shaker) at room temperature for 15 min.
13. Remove the samples from chips.
    a. Add 200 µL of PBS (Wash 1) to chips; or
    b. Add 200 µL of 1 M urea 0.125% CHAPS, 0.5 M NaCl, 100 mM acetate, pH 4.5 (Wash 2) to chips.
14. Mix vigorously (250 rpm on a platform shaker) at room temperature for 5 min.
15. Remove washes from spots.
    a. Add fresh 200 µL of PBS (Wash 1) to chips; or.
    b. Add fresh 200 µL of 1 M urea, 0.125% CHAPS, 0.5 M NaCl,

100 mM acetate, pH 4.5 (Wash 2) to chips.

16. Mix vigorously (250 rpm on a platform shaker) at room temperature for 5 min.
17. Repeat buffer wash one more time.
18. Wash chips 3 times with water by filling the wells and emptying.
19. Remove chips from the Bioprocessor.
20. Rinse chips with deionized water.
21. Air dry chips.
22. Add 0.5 µL of SPA or CHCA to chips two times (air dry spots between additions).

VI. PRE-FRACTIONATION AND GAS PHASE ION SPECTROMETRY ANALYSIS, AND DETECTED MARKERS

A. THE SINGLE STRANDED DNA PROTOCOL, AND MARKERS DETECTED

1. Protocols and Sample Preparation

A "single stranded DNA protocol" refers to a pre-fractionation protocol which comprises applying a sample to a Cibacron blue agarose column to remove serum albumin present in the sample followed by applying the sample to single stranded DNA agarose beads in 96 well filter plates and sequentially eluting the sample. A schematic diagram of "a single stranded DNA protocol" is shown in Figure 1. However, other permutations of the single stranded DNA protocol may be used to extract different markers at each level.

1. To 40 µL human serum add 60 µL 8M urea (*e.g.*, Sigma electrophoresis grade), 1% CHAPS in PBS.
2. Vortex in cold, 10 min.
3. Prepare Cibacron blue agarose spin column. Alternatively, prepare anti-HSA antibody coupled beads. Cibacron Blue agarose and anti-HSA can be purchased from Sigma. Sheep anti-human serum albumin antibody was also obtained from Biodesign, Saco, Maine. The antibodies can be covalently coupled to preactivated beads (*e.g.*, from Pierce) and packed into column. Start with 150 µL of a 50% (v/v) Cibacron Blue Agarose

bead suspension. Equilibrate with 3 x 300 µL 1M urea, 0.125% CHAPS, PBS. Spin column dry at 1000 x g, 30 sec each time.

4. Add 100 µL diluted serum sample to column.
5. Wash original serum tube with 100 µL 1M urea, 0.125% CHAPS, PBS. Add wash to column.
6. Put column into fresh collection tube. Vortex in cold, 15 min.
7. Centrifuge column at 1000 x g, 30 sec. Save filtrate in the collection tube.
8. Add 100 µL 1 M urea, 0.125% CHAPS, PBS to the column. Note the decreased amount of urea and CHAPS in this buffer.
9. Put column into fresh collection tube. Vortex in cold, 15 min.
10. Put column back into first collection tube. Spin column 1000 x g, 30 sec.
11. Pool the 2 filtrates. Volume should equal approximately 300 µL.
12. Single-stranded DNA agarose bead (GIBCO BRL) was washed and stored as a 50% suspension (v/v) in 50mM Tris HCl pH8.
13. 150µL of ss-DNA agarose suspension was added to each well of a 96-well filter plate (Lyprodyne membrane, 0.45µm pore size) and was placed on the vacuum manifold of a Biomek 2000 Laboratory Automation Workstation (Beckman Coulter, CA). All liquid handling and filtration were performed on the Workstation. Wells containing agarose beads were washed four times, each with 150µL of 1M urea 0.1% CHAPS 50mM Tris HCl pH8, and vacuum-suctioned dry.
14. 25µL of albumin-depleted human serum sample were mixed with 100µL of 1M urea 0.1% CHAPS 50mM Tris HCl pH8 in a 96-well microtiter plate and applied to one well of ss-DNA column in the 96-well filter plate.
15. Column beads were mixed with serum sample on a programmable reciprocating shaking platform (MicroMix5, DPC Cirrus Inc, NJ) at 4°C for 15 minutes.
16. Filter plate was returned to vacuum manifold. Flowthrough fraction was collected by applying vacuum and into a 96-well microtiter plate (plate D1).
17. 100µL of 1M urea 0.1% CHAPS 50mM Tris HCl pH8 were added to each well of ss-DNA agarose column. Agarose beads were mixed at 4°C for 5

minutes. Wash solution was collected on the vacuum manifold and combined with flowthrough fraction in plate D1.

18. ss-DNA agarose beads were washed similarly two more times, each with 100µL of 50mM Tris HCl 1M sodium chloride 0.1% OGP (n-octyl-β-D-glucopyranoside) pH7.2. Both washes were collected and combined in a microtiter plate (plate D2).

19. Process each fraction, (D1 and D2) on a set of chips as described below.

a) Protein Profiling on an IMAC3 copper ProteinChip® array

C1. IMAC3 ProteinChip® arrays were set up in a 96-well bioprocessor (Ciphergen Biosystems Inc.) and placed on a programmable reciprocating shaking platform (MicroMix5, DPC Cirrus Inc, NJ) integrated with a Biomek 2000 Laboratory Automation Workstation (Beckman Coulter, CA). All liquid handling and mixing were performed on the Workstation.

C2. 50µL of 0.1M copper sulfate in water solution were loaded to each well of 96-well bioprocessor for 5 minutes. After 10 minute of shaking on the shaking platform, the copper solution was removed and unbound copper was washed off with 300 µL water two times and 200 µL 0.1 M sodium acetate pH4 buffer

C3. Each array was equilibrated two times, each with 150µL of Phosphate Buffered Saline (PBS) pH7.2 for 5 minutes.

C4. 20µL of each serum fraction collected after passing through ss-DNA agarose column was mixed with 60µL of PBS pH7.2 in each well of 96-well bioprocessor. Incubation was performed on the platform shaker at room temperature for 20 minutes.

C5. ProteinChip® arrays were washed three times, each with 150µL of PBS pH7.2 for 5 minutes.

C6. Arrays were rinsed with water and air-dried. 0.5µL of sinapinic acid solution (in 50% acetonitrile and 0.5% trifluoroacetic acid) was applied to each array two times and air-dried. Sinapinic acid solution was air-dried between additions.

C7. Arrays were analyzed in a ProteinChip® reader (Ciphergen Biosystems Inc.) and spectra from 130 laser shots were averaged.

b) Protein Profiling on a WCX2 ProteinChip® array.

W1. WCX2 ProteinChip® arrays were set up in a 96-well bioprocessor (Ciphergen Biosystems Inc.) and placed on a programmable reciprocating shaking platform (MicroMix5, DPC Cirrus Inc, NJ) integrated with a Biomek 2000 Laboratory Automation Workstation (Beckman Coulter, CA). All liquid handling and mixing were performed on the Workstation.

W2. Each array was equilibrated two times, each with 150μL of 0.1M sodium acetate pH4 buffer for 5 minutes.

W3. 20μL of each serum fraction collected after passing through ss-DNA agarose column was mixed with 60μL of 0.1M sodium acetate pH4 buffer in each well of 96-well bioprocessor. Incubation was performed on the platform shaker at room temperature for 20 minutes.

W4. ProteinChip® arrays were washed three times, each with 150μL of 0.1M sodium acetate pH4 buffer for 5 minutes.

W5. Arrays were rinsed with water and air-dried. 0.5μL sinapinic acid solution (in 50% acetonitrile and 0.5% trifluoroacetic acid) was applied to each array two times and air-dried. Sinapinic acid solution was air-dried between additions.

W6. Arrays were analyzed in a ProteinChip® reader (Ciphergen Biosystems Inc.) and spectra from 130 laser shots were averaged.

c) Protein Profiling on a SAX2 ProteinChip® array.

S1. SAX2 ProteinChip® arrays were set up in a 96-well bioprocessor (Ciphergen Biosystems Inc.) and placed on a programmable reciprocating shaking platform (MicroMix5, DPC Cirrus Inc, NJ) integrated with a Biomek 2000 Laboratory Automation Workstation (Beckman Coulter, CA). All liquid handling and mixing were performed on the Workstation.

S2. Each array was equilibrated two times, each with 150μL of 50mM Tris HCl pH8 for 5 minutes.

S3. 20μL of each serum fraction collected after passing through ss-DNA agarose column was mixed with 60μL of 50mM Tris HCl pH8 in each well of 96-well bioprocessor. Incubation was performed on the platform shaker at room temperature for 20 minutes.

S4. ProteinChip® arrays were washed three times, each with 150μL of 50mM Tris HCl pH8 for 5 minutes.

S5. Arrays were rinsed with water and air-dried. 0.5μL sinapinic acid solution (in 50% acetonitrile and 0.5% trifluoroacetic acid) was applied to each array two times and air-dried. Sinapinic acid solution was air-dried between additions.

S6. Arrays were analyzed in a ProteinChip reader (Ciphergen Biosystems Inc.) and spectra from 130 laser shots were averaged.

2. Markers Detected

18 serum samples from breast cancer patients (Stage IV) and 18 serum samples from control women (*i.e.*, women with undetectable breast cancer) were prepared as described above. The prepared samples were analyzed using the SAX2 chip, the IMAC3-Cu^{++} chip, and the WCX2 chip. One of skill in the art would realize that a number of other chips with different surface chemistries can be used to extract different proteins.

The following markers were detected from the above samples: markers Br 1 and Br 2 were detected in fraction D1 and on a SAX2 chip; markers Br 3-Br 5 were detected in fraction D1 and on a Cu(II) chip; markers Br 6-13 were detected in fraction D1 and on a WCX2 chip, marker Br 14 was detected in fraction D2 and on a Cu(II) chip; markers Br 15-Br 19 were detected in fraction D2 and on a WCX2 chip. Certain markers were better resolved using SPA, while other markers were better resolved using CHCA as an energy absorbing material.

B. THE ANION EXCHANGE/HEPARIN SPIN COLUMN PROTOCOL AND MARKERS DETECTED

1. Protocol and Sample Preparation

An "anion exchange/heparin protocol" refers to a pre-fractionation protocol which comprises applying a sample to anion exchange beads in a 96 well filter plate and sequentially eluting the sample with buffers of varying pH. The eluates are further fractionated by applying the eluates to heparin beads in a 96 well filter plate, and eluted with buffers of varying ionic strength. The schematic diagram of the anion exchange/heparin spin column protocol is shown in Figure 2. However, other

permutations of the anion exchange/heparin protocol may be used to extract different markers at each level.

1. To 20 µL human serum, add 30 µL 8 M urea, 1% CHAPS, 0.1M Tris buffer pH 9.
2. Q HyperD F beads (Biosepra, France) were washed and stored in a 50% suspension (v/v) 100 mM sodium Bicarbonate pH 8.2.
3. 150 µl Q HyperD F anion exchange beads were packed in each well of a 96 well filter plate (Loprodyne membrane, 0.45µm pore size) and was placed on a vacuum manifold of a Biomek 2000 Laboratory Automation Workstation (Beckman Coulter, CA). All liquid handling and filtration were performed on the workstation. Beads were washed with three times each with 200 µl of 100 mM sodium bicarbonate, pH 8.2 and vacuum suctioned dry.
4. The entire diluted human serum sample was added to a well containing the Q HyperD F beads and 100 µl of 100 mM soduim bicarbonate buffer, pH 8.2 and mixed on a programmable reciprocating shaking platform (MicroMix5, DPC Cirrus, Inc., NJ) at 4^0C for 15 minutes
5. The filter plate was returned to the vacuum manifold, the flowthrough fraction was collected by applying vacuum and collecting the fluid in a 96 well collection plate Q1.
6. 100 µl of 100 mM ammonium acetate, pH 7 was added to each well of the Q column and returned to the platform shaker for mixing at 4^0C for 15 minutes.
7. The filter plate was returned to the vacuum manifold and buffer 1 was collected and combined with the Q1 fraction.
8. 100 µl of 100 mM sodium acetate, pH 5, was added to each well of the Q column, mixed as described above and sample was collected creating fraction Q2. This wash was repeated once, pooling into fraction Q2.
9. 100 µl of 100 mM sodium citrate, pH 3 was added to each well of the Q column, mixed as described and sample was collected on the vacuum manifold creating fraction Q3. This wash was repeated once, pooling into fraction Q3.

10. 100 µl of buffer 4 containing 33.3% isopropanol, 16.7% acetonitrile, 0.1% trifluroacetic acid (TFA) was added to each well of the Q column, mixed as described and collected on the vacuum manifold creating fraction Q4. This wash was repeated once, pooling into fraction Q4.

11. Fractions Q1 and Q2 were combined and fractions Q3 and Q4 were combined and further fractionated using a heparin bead (Heparin HyperD M, Biosepra, France) 96 well filtration plate (Loprodyne membrane, 0.45µm pore size).

12. Heparin HyperD M bead (Biosepra) was washed and stored as a 50% suspension (v/v) in Phosphate Buffered Saline (PBS) pH7.2.

13. 150µL of Heparin HyperD M suspension was added to each well of a 96-well filter plate (Loprodyne membrane, 0.45µm pore size) and was placed on the vacuum manifold of a Biomek 2000 Laboratory Automation Workstation (Beckman Coulter, CA). All liquid handling and filtration were performed on the Workstation. Column was washed three times, each with 200µL of 1M urea 0.1% CHAPS PBS pH7.2, and vacuum-suctioned dry.

14. 25µL of fraction 1 and 25µL of fraction 2 collected after passing through Q HyperD F column were combined with 75µL of 2M urea 0.25% CHAPS PBS pH7.2 in a 96-well microtiter plate, and the mixture was applied to one well of Heparin HyperD M column in the 96-well filter plate.

15. 20µL of fraction 3 and 20µL of fraction 4 collected after passing through Q HyperD F column were combined with 75µL of 2M urea 0.5% CHAPS 50mM Tris HCl pH9 buffer in a 96-well microtiter plate, and the mixture was applied to one well of Heparin HyperD M column in the 96-well filter plate.

16. Column beads were mixed with serum fractions on a programmable reciprocating shaking platform (MicroMix5, DPC Cirrus Inc, NJ) at 4°C for 15 minutes.

17. Filter plate was returned to vacuum manifold. Flowthrough fraction was collected by applying vacuum and into a 96-well microtiter plate (plate H1).

18. 100µL of 1M urea 0.1% CHAPS PBS pH7.2 were added to each well of Heparin column. Column beads were mixed on platform shaker at 4°C for 15 minutes. Wash solution was collected on the vacuum manifold and combined with flowthrough fraction in plate H1.

19. Heparin columns were washed similarly two more times, each with 100µL of 100mM Tris HCl 2.5M sodium chloride 0.1% OGP (n-octyl-β-D-glucopyranoside) pH7.2. Both washes were collected and combined in a microtiter plate (plate H2).

20. Heparin columns were washed one time with 100µL of 1.5M guanidine thiocyanate 0.25% OGP 50mM HEPES pH7, followed by 100µL of 33.3% isopropanol, 16.7% acetonitrile, 0.1% trifluoroacetic acid in water. Both washes were combined and collected in a microtiter plate (plate H3).

a) Protein Profiling on IMAC3 copper ProteinChip® array

C1. IMAC3 ProteinChip® arrays were set up in a 96-well bioprocessor (Ciphergen Biosystems Inc.) and placed on a programmable reciprocating shaking platform (MicroMix5, DPC Cirrus Inc, NJ) integrated with a Biomek 2000 Laboratory Automation Workstation (Beckman Coulter, CA). All liquid handling and mixing were performed on the Workstation.

C2. 50µL of 0.1M copper sulfate in water solution were loaded to each well of 96-well bioprocessor for 5 minutes. After 10 minute of shaking on the shaking platform, the copper solution was removed and unbound copper was washed off with 300 µL water two times and 200 µL 0.1 M sodium acetate pH4 buffer.

C3. Each array was equilibrated two times, each with 200µL of 50mM HEPES pH7 for 5 minutes.

C4. 50µL of each serum fraction collected after passing through Heparin HyperD M column were mixed with 150µL of 50mM HEPES pH7 in each well of 96-well bioprocessor. Incubation was performed on the platform shaker at room temperature for 20 minutes.

C5. ProteinChip® arrays were washed three times, each with 200µL of 50mM HEPES pH7 for 5 minutes.

C6. Arrays were rinsed with water and air-dried. 0.5μL of sinapinic acid solution (in 50% acetonitrile and 0.5% trifluoroacetic acid) was applied to each array two times and air-dried. Sinapinic acid solution was air-dried between additions.

C7. Arrays were analyzed in a ProteinChip® reader (Ciphergen Biosystems Inc.) and spectra from 130 laser shots were averaged.

b) SAX2 and H4 ProteinChip arrays

S1. SAX2 and H4 ProteinChip® arrays were set up in a 96-well bioprocessor (Ciphergen Biosystems Inc.) and placed on a programmable reciprocating shaking platform (MicroMix5, DPC Cirrus Inc, NJ) integrated with a Biomek 2000 Laboratory Automation Workstation (Beckman Coulter, CA). All liquid handling and mixing were performed on the Workstation.

S2. Each array was equilibrated two times, each with 200μL of 50mM HEPES pH7 for 5 minutes.

S3. 50μL of each serum fraction collected after passing through Heparin HyperD M column were mixed with 150μL of 50mM HEPES pH7 in each well of 96-well bioprocessor. Incubation was performed on the platform shaker at room temperature for 20 minutes.

S4. ProteinChip® arrays were washed three times, each with 200μL of 50mM HEPES pH7 for 5 minutes.

S5. Arrays were rinsed with water and air-dried. 0.5μL sinapinic acid solution (in 50% acetonitrile and 0.5% trifluoroacetic acid) was applied to each array two times and air-dried. Sinapinic acid solution was air-dried between additions.

S6. Arrays were analyzed in a ProteinChip® reader (Ciphergen Biosystems Inc.) and spectra from 130 laser shots were averaged.

2. Markers Detected

18 serum samples from breast cancer patients (Stage IV) and 18 serum samples from control women (*i.e.*, women with undetectable breast cancer) were prepared as described above. The prepared samples were analyzed using the SAX2 chip, the IMAC3-Cu^{++} chip, and the WCX2 chip. One of skill in the art would realize that a

number of other chips with different surface chemistries can be used to extract different proteins.

The following markers were detected from the above samples: markers Br 20 and 21 were detected in Q3/Q4/H1 fractions and on the Cu(II) chip; markers Br 22-Br 29 were detected in Q3/Q4/H1 fractions and on the SAX2 chip; markers Br 30-Br 33 were detected in Q3/Q4/H2 fractions and on the Cu(II) chip. Certain markers were better resolved using SPA, while other markers were better resolved using CHCA as an energy absorbing material.

C. THE LECTIN PROTOCOL AND MARKERS DETECTED

1. Protocol and Sample Preparation.

A "lectin protocol" refers to a pre-fractionation protocol which comprises applying a sample to an anti-human serum albumin column to remove serum albumin present in the sample followed by applying the sample to lectin beads in a 96 well filter plate and sequentially eluting the sample with buffers with varying concentrations of salt, sugars and pH. The schematic diagram of a lectin protocol is shown in Figure 3. However, other permutations of the lectin protocol may be used to extract different markers at each level.

1. To 20 µL human serum add 30 µL 8M urea (*e.g.*, Sigma electrophoresis grade), 1% CHAPS in PBS.
2. Vortex in cold, 10 min.
3. Prepare anti-HSA antibody coupled beads. Sheep anti-human serum albumin antibody was also obtained from Biodesign, Saco, Maine. The antibodies were immobilized on Protein G Sepharose (Pharmacia) or can be covalently coupled to preactivated beads (*e.g.*, from Pierce) and packed into column. Start with 250 µL of a 50% (v/v) anti-HSA Protein G bead suspension. Equilibrate with 3 x 300 µL PBS. Spin column dry at 1000 x g, 30 sec each time.
4. Add 15 µL diluted serum sample to 100 µL PBS in anti-HSA column.
5. Put column into fresh collection tube. Vortex in cold, 90 min.
6. Centrifuge column at 1000 x g, 30 sec. Save filtrate in the collection tube.
7. Add 100 µL 1 M urea, 0.125% CHAPS, PBS to the column.
8. Put column into fresh collection tube. Vortex in cold, 15 min.
9. Put column back into first collection tube. Spin column 1000 x g, 30 sec.

10. Pool the 2 filtrates.

11. Repeat wash. Final volume should equal approximately 300 µL.

12. Mix 50 µl of albumin depleted serum sample with 100 µl of 0.1 M Tris HCl, 0.1% OGP, 0.5 M NaCl.

13. Wheat Germ Lectin agarose beads (Sigma Chemical, MO) were washed and stored as a 50% suspension (v/v) in 0.1M Tris HCl 0.1% OGP (n-octyl-β-D-glucopyranoside) 0.5M NaCl pH7.2.

14. 100µL of Wheat Germ Lectin agarose suspension was added to each well of a 96-well filter plate (Loprodyne membrane, 0.45µm pore size) and was placed on the vacuum manifold of a Biomek 2000 Laboratoratory Automation Workstation (Beckman Coulter, CA). All liquid handling and filtration were performed on the Workstation. Columns were washed four times, each with 200µL of 0.1M Tris HCl 0.1% OGP 0.5M NaCl pH7.2, and vacuum-suctioned dry.

15. 50µL of albumin-depleted human serum were combined with 100µL of 0.1M Tris HCl 0.1% OGP 0.5M NaCl pH7.2 in a 96-well microtiter plate, and the mixture was applied to one well of Wheat Germ Lectin agarose column in the 96-well filter plate.

16. Column beads were mixed with serum on a programmable reciprocating shaking platform (MicroMix5, DPC Cirrus Inc, NJ) at 4°C for 15 minutes.

17. Filter plate was returned to vacuum manifold. Flowthrough fraction was collected by applying vacuum and into a 96-well microtiter plate (plate W1).

18. 70µL of 0.1M Tris HCl 0.1% OGP 0.5M NaCl pH7.2 were added to each column. Column beads were mixed on platform shaker at 4°C for 3 minutes. Wash solution was collected on the vacuum manifold and combined with flowthrough fraction in plate W1.

19. Lectin columns were washed similarly three more times, each with 70µL of 0.1M Tris HCl 0.1% OGP 0.5M NaCl pH7.2. Washes were collected and combined in microtiter plate W1.

20. Lectin columns were washed similarly three times, each with 70µL of 25mM N-acetyl glucosamine 0.1% OGP PBS pH7.2. Washes were collected and combined in a microtiter plate (plate W2).

21. Lectin columns were washed similarly two times, each with 70µL of 100mM sodium borate 0.1% OGP pH6.5. Washes were collected and combined in a microtiter plate (plate W3). Finally, columns were washed one time for 6 minutes with 70µL of 100mM citrate phosphate 0.1% OGP pH3 and the wash solution was collected and added to plate W3.

a) Protein Profiling on an IMAC3 copper ProteinChip® array

C1. IMAC3 ProteinChip® arrays were set up in a 96-well bioprocessor (Ciphergen Biosystems Inc.) and placed on a programmable reciprocating shaking platform (MicroMix5, DPC Cirrus Inc, NJ) integrated with a Biomek 2000 Laboratory Automation Workstation (Beckman Coulter, CA). All liquid handling and mixing were performed on the Workstation.

C2. 50µL of 0.1M copper sulfate in water solution were loaded to each well of 96-well bioprocessor for 5 minutes. After 10 minute of shaking on the shaking platform, the copper solution was removed and unbound copper was washed off with 300 µL water two times and 200 µL 0.1 M sodium acetate pH4 buffer.

C3. Each array was equilibrated two times, each with 150µL of Phosphate Buffered Saline (PBS) pH7.2 for 5 minutes.

C4. 20µL of each serum fraction collected after passing through Wheat Germ Lectin agarose column were mixed with 40µL of PBS pH7.2 in each well of 96-well bioprocessor. Incubation was performed on the platform shaker at room temperature for 20 minutes.

C5. ProteinChip® arrays were washed three times, each with 150µL of PBS pH7.2 for 5 minutes.

C6. Arrays were rinsed with water and air-dried. 0.5µL of sinapinic acid solution (in 50% acetonitrile and 0.5% trifluoroacetic acid) was applied to each array two times and air-dried. Sinapinic acid solution was air-dried between additions.

C7. Arrays were analyzed in a ProteinChip® reader (Ciphergen Biosystems Inc.) and spectra from 130 laser shots were averaged.

b) Protein Profiling on a WCX2 ProteinChip® array

W1. WCX2 ProteinChip® arrays were set up in a 96-well bioprocessor (Ciphergen Biosystems Inc.) and placed on a programmable reciprocating shaking platform (MicroMix5, DPC Cirrus Inc, NJ) integrated with a Biomek 2000 Laboratory Automation Workstation (Beckman Coulter, CA). All liquid handling and mixing were performed on the Workstation.

W2. Each array was equilibrated two times, each with 150μL of 0.1M sodium phosphate pH6 buffer for 5 minutes.

W3. 20μL of each serum fraction collected after passing through Wheat Germ Lectin agarose column was mixed with 40μL of 0.1M sodium phosphate pH6 buffer in each well of 96-well bioprocessor. Incubation was performed on the platform shaker at room temperature for 20 minutes.

W4. ProteinChip® arrays were washed three times, each with 150μL of 0.1M sodium phosphate pH6 buffer for 5 minutes.

W5. Arrays were rinsed with water and air-dried. 0.5μL sinapinic acid solution (in 50% acetonitrile and 0.5% trifluoroacetic acid) was applied to each array two times and air-dried. Sinapinic acid solution was air-dried between additions.

W6. Arrays were analyzed in a ProteinChip® reader (Ciphergen Biosystems Inc.) and spectra from 130 laser shots were averaged.

c) Protein Profiling on a SAX2 ProteinChip® array

S1. SAX2 ProteinChip® arrays were set up in a 96-well bioprocessor (Ciphergen Biosystems Inc.) and placed on a programmable reciprocating shaking platform (MicroMix5, DPC Cirrus Inc, NJ) integrated with a Biomek 2000 Laboratory Automation Workstation (Beckman Coulter, CA). All liquid handling and mixing were performed on the Workstation.

S2. Each array was equilibrated two times, each with 150μL of 50mM Tris HCl pH8 for 5 minutes.

S3. 20μL of each serum fraction collected after passing through Wheat Germ Lectin column was mixed with 40μL of 50mM Tris HCl pH8 in each well of 96-well bioprocessor. Incubation was performed on the platform shaker at room temperature for 20 minutes.

S4. ProteinChip® arrays were washed three times, each with 150μL of 50mM Tris HCl pH8 for 5 minutes.

S5. Arrays were rinsed with water and air-dried. 0.5μL sinapinic acid solution (in 50% acetonitrile and 0.5% trifluoroacetic acid) was applied to each array two times and air-dried. Sinapinic acid solution was air-dried between additions.

S6. Arrays were analyzed in a ProteinChip® reader (Ciphergen Biosystems Inc.) and spectra from 130 laser shots were averaged.

2. Markers Detected

Using the lectin protocol described above, the following markers were detected: markers Br 34 and Br 35 were detected in W1 and on the SAX2 chip; and markers Br 36 and Br 37 were detected in W2 and on the Cu(II) chip. One of skill in the art would realize that a number of other chips with different surface chemistries can be used to extract different proteins. Certain markers were better resolved using SPA, while other markers were better resolved using CHCA as an energy absorbing material.

The present invention provides novel materials and methods for aiding breast cancer diagnosis using markers that are differentially present in samples of a breast cancer patient and a normal subject who does not have breast cancer. While specific examples have been provided, the above description is illustrative and not restrictive. Any one or more of the features of the previously described embodiments can be combined in any manner with one or more features of any other embodiments in the present invention. Furthermore, many variations of the invention will become apparent to those skilled in the art upon review of the specification. The scope of the invention should, therefore, be determined not with reference to the above description, but instead should be determined with reference to the appended claims along with their full scope of equivalents.

All publications and patent documents cited in this application are incorporated by reference in their entirety for all purposes to the same extent as if each individual publication or patent document were so individually denoted. By their citation of various references in this document, Applicants do not admit any particular reference is "prior art" to their invention.

WHAT IS CLAIMED IS:

1. A method for aiding a breast cancer diagnosis, the method comprising:

(a) detecting at least one protein marker in a sample, wherein the protein marker is selected from:

Marker Br 1: 4170 ± 8 Da;
Marker Br 2: 21080 ± 42 Da;
Marker Br 3: 9339 ± 19 Da;
Marker Br 4: 28308 ± 57 Da;
Marker Br 5: 28344 ± 57 Da;
Marker Br 6: 4148 ± 9 Da;
Marker Br 7: 5634 ± 11 Da;
Marker Br 8: 6520 ± 13 Da;
Marker Br 9: 6955 ± 14 Da;
Marker Br 10: 7507 ± 15 Da;
Marker Br 11: 9116 ± 16 Da;
Marker Br 12: 9453 ± 19 Da;
Marker Br 13: 17310 ± 35 Da;
Marker Br 14: 89805 ± 449 Da;
Marker Br 15: 4256 ± 9 Da;
Marker Br 16: 4357 ± 9 Da;
Marker Br 17: 4470 ± 9 Da;
Marker Br 18: 9292 ± 19 Da;
Marker Br 19: 9335 ± 19 Da;
Marker Br 20: 11758 ± 24 Da;
Marker Br 21: 11776 ± 24 Da;
Marker Br 22: 13913 ± 28 Da;
Marker Br 23: 17291 ± 35 Da;
Marker Br 24: 17419 ± 35 Da;
Marker Br 25: 21103 ± 42 Da;
Marker Br 26: 40297 ± 81 Da;
Marker Br 27: 4488 ± 9 Da;
Marker Br 28: 4647 ± 9 Da;

Marker Br 29: 80428 ± 402 Da;

Marker Br 30: 11757 ± 24 Da;

Marker Br 31: 4487 ± 9 Da;

Marker Br 32: 5360 ± 11 Da;

Marker Br 33: 11773 ± 24 Da;

Marker Br 34: 13977 ± 28 Da;

Marker Br 35: 14004 ± 28 Da;

Marker Br 36: 51700 ± 259 Da; and

Marker Br 37: 153894 ± 769 Da; and

(b) correlating the detection of the marker or markers with a probable diagnosis of breast cancer.

2. The method of claim 1, wherein the correlation takes into account the amount of the marker or markers in the sample compared to a control amount of the marker or markers.

3. The method of claim 1, wherein the correlation takes into account the presence or absence of the marker or markers in the sample and the frequency of detection of the same marker or markers in a control.

4. The method of claim 2, wherein the correlation further takes into account the presence or absence of the marker or markers in the sample and the frequency of detection of the same marker or markers in a control.

5. The method of claim 1, wherein the method comprises detecting a plurality of the markers.

6. The method of claim 1, wherein the method comprises detecting at least three of the markers.

7. The method of claim 1, wherein the sample is blood serum.

8. The method of claim 1, wherein gas phase ion spectrometry is used for detecting the marker or markers.

9. The method of claim 8, wherein the gas phase ion spectrometry is laser desorption/ionization mass spectrometry.

10. The method of claim 9, wherein laser desorption/ionization mass spectrometry comprises:

(a) providing a substrate comprising an adsorbent attached thereto;

(b) contacting the sample with the adsorbent; and

(c) desorbing and ionizing the marker or markers from the substrate and detecting the desorbed/ionized marker or markers with the mass spectrometer.

11. The method of claim 10, wherein the substrate is a probe adapted for use with the mass spectrometer.

12. The method of claim 10, wherein the substrate is suitable for being placed on a probe which is adapted for use with the mass spectrometer.

13. The method of claim 10, wherein the adsorbent is an antibody that specifically binds to the marker.

14. The method of claim 10, wherein the adsorbent is a cationic adsorbent.

15. The method of claim 14, comprising detecting two or more of markers Br 1, Br 2, Br 22, Br 23, Br 24, Br 25, Br 26, Br 27, Br 28, Br 29, Br 34 and Br 35.

16. The method of claim 10, wherein the adsorbent is a metal chelating adsorbent.

17. The method of claim 16, comprising detecting two or more of markers Br 3, Br, 4, Br 5, Br 14, Br 20, Br 21, Br 30, Br 31, Br 32, Br 33, Br 36 and Br 37.

18. The method of claim 10, wherein the adsorbent is an anionic adsorbent.

19. The method of claim 18, comprising detecting two or more of the markers Br 6, Br 7, Br 8, Br 9, Br 10, Br 11, Br 12, Br 13, Br 15, Br 16, Br 17, Br 18, and Br 19.

20. The method of claim 1, wherein an immunoassay is used for detecting the marker or markers.

21. The method of claim 9, the method further comprising:

(a) generating data on the sample with the mass spectrometer indicating intensity of signal for mass/charge ratios;

(b) transforming the data into computer-readable form; and

(c) operating a computer to execute an algorithm, wherein the algorithm determines closeness-of-fit between the computer-readable data and data indicating a diagnosis of breast cancer or a negative diagnosis.

22. A method for detecting at least one protein marker in a sample, wherein the marker is selected from:

Marker Br 1: 4170 ± 8 Da;
Marker Br 2: 21080 ± 42 Da;
Marker Br 3: 9339 ± 19 Da;
Marker Br 4: 28308 ± 57 Da;
Marker Br 5: 28344 ± 57 Da;
Marker Br 6: 4148 ± 9 Da;
Marker Br 7: 5634 ± 11 Da;
Marker Br 8: 6520 ± 13 Da;
Marker Br 9: 6955 ± 14 Da;
Marker Br 10: 7507 ± 15 Da;
Marker Br 11: 9116 ± 16 Da;
Marker Br 12: 9453 ± 19 Da;
Marker Br 13: 17310 ± 35 Da;
Marker Br 14: 89805 ± 449 Da;
Marker Br 15: 4256 ± 9 Da;
Marker Br 16: 4357 ± 9 Da;
Marker Br 17: 4470 ± 9 Da;
Marker Br 18: 9292 ± 19 Da;
Marker Br 19: 9335 ± 19 Da;
Marker Br 20: 11758 ± 24 Da;
Marker Br 21: 11776 ± 24 Da;

Marker Br 22: 13913 ± 28 Da;
Marker Br 23: 17291 ± 35 Da;
Marker Br 24: 17418 ± 35 Da;
Marker Br 25: 21103 ± 42 Da;
Marker Br 26: 40297 ± 81 Da;
Marker Br 27: 4487 ± 9 Da;
Marker Br 28: 4646 ± 9 Da;
Marker Br 29: 80428 ± 402 Da;
Marker Br 30: 11757 ± 24 Da;
Marker Br 31: 4487 ± 9 Da;
Marker Br 32: 5360 ± 11 Da;
Marker Br 33: 11773 ± 24 Da;
Marker Br 34: 13977 ± 28 Da;
Marker Br 35: 14004 ± 28 Da;
Marker Br 36: 51700 ± 259 Da; and
Marker Br 37: 153894 ± 769 Da;

wherein the method comprises detecting the marker or markers by gas phase ion spectrometry.

23. The method of claim 22, wherein the sample is a blood serum sample.

24. The method of claim 22, wherein the detection method comprises detecting the marker or markers by laser desorption/ionization mass spectrometry.

25. The method of claim 22, further comprising comparing the amount of the detected marker or markers to a control.

26. The method of claim 24 comprising:

(a) generating data on the sample with the mass spectrometer indicating intensity of signal for mass/charge ratio;

(b) transforming the data into computer-readable form; and

(c) operating a computer and executing an algorithm that detects signal in the computer-readable data representing the marker or markers.

27. The method of claim 24, wherein laser desorption/ionization mass spectrometry comprises:

(a) providing a substrate comprising an adsorbent attached thereto;

(b) contacting the sample with the adsorbent; and

(c) desorbing and ionizing the marker or markers from the substrate and detecting the desorbed/ionized marker or markers with the mass spectrometer.

28. The method of claim 27, wherein the substrate is a probe adapted for use with the mass spectrometer.

29. The method of claim 27, wherein the substrate is suitable for being placed on a probe which is adapted for use with the mass spectrometer.

30. The method of claim 24 comprising:

(a) removing serum albumin from the sample;

(b) binding the markers in the sample to ssDNA, and eluting the marker or markers into a sample fraction;

(b) contacting the sample fraction with an adsorbent on a substrate;

(c) desorbing and ionizing the marker or markers retained on the adsorbent and detecting the desorbed/ionized marker or markers with the mass spectrometer.

31. The method of claim 30, wherein the adsorbent is selected from the group consisting of a cationic adsorbent, an anionic adsorbent, or a metal chelating adsorbent.

32. The method of claim 30, comprising detecting two or more of markers Br 1, Br 2, Br 3, Br 4, Br, 5, Br 6, Br 7, Br 8, Br 9, Br 10, Br 11, Br 12, Br 13, Br 14, Br 15, Br 16, Br 17, Br 18, and Br 19.

33. The method of claim 24 comprising:

(a) removing serum albumin from the sample;

(b) passing the marker or markers in the sample over a lectin column;

(c) eluting the marker or markers from the lectin column into a sample fraction;

(d) contacting the sample fraction with an adsorbent on a substrate;

(e) desorbing and ionizing marker or markers retained on the adsorbent and detecting the desorbed/ionized marker or markers with the mass spectrometer.

34. The method of claim 33, wherein the adsorbent is selected from the group consisting of a cationic adsorbent, an anionic adsorbent, or a metal chelate adsorbent.

35. The method of claim 34, comprising detecting two or more of markers Br 34, Br 35, Br 36, and Br 37.

36. The method of claim 24, the method comprising:

(a) fractionating the sample by anion exchange chromatography and collecting a first sample fraction containing the marker or markers;

(b) fractionating sample fraction 1 on a heparin column and collecting a second sample fraction containing the marker or markers;

(c) contacting the second sample fraction with an adsorbent on a substrate;

(c) desorbing and ionizing marker or markers retained on the adsorbent and detecting the desorbed/ionized marker or markers with the mass spectrometer.

37. The method of claim 36, wherein the adsorbent is selected from the group consisting of a cationic adsorbent, an anionic adsorbent, or a metal chelate adsorbent.

38. The method of claim 37, comprising detecting two or more of the markers Br 20, Br 21, Br 22, Br 23, Br 24, Br 25, Br 26, Br 27, Br 28, Br 29, Br 30, Br 31, Br 32, and Br 33.

39. A method for detecting at least one protein marker in a sample, wherein the marker is selected from:

Marker Br 1: 4170 ± 8 Da;

Marker Br 2: 21080 ± 42 Da;

Marker Br 3: 9339 ± 19 Da;

Marker Br 4: 28308 ± 57 Da;

Marker Br 5: 28344 ± 57 Da;

Marker Br 6: 4148 ± 9 Da;

Marker Br 7: 5634 ± 11 Da;

Marker Br 8: 6520 ± 13 Da;
Marker Br 9: 6955 ± 14 Da;
Marker Br 10: 7507 ± 15 Da;
Marker Br 11: 9116 ± 16 Da;
Marker Br 12: 9453 ± 19 Da;
Marker Br 13: 17310 ± 35 Da;
Marker Br 14: 89805 ± 449 Da;
Marker Br 15: 4256 ± 9 Da;
Marker Br 16: 4357 ± 9 Da;
Marker Br 17: 4470 ± 9 Da;
Marker Br 18: 9292 ± 19 Da;
Marker Br 19: 9335 ± 19 Da;
Marker Br 20: 11758 ± 24 Da;
Marker Br 21: 11776 ± 24 Da;
Marker Br 22: 13913 ± 28 Da;
Marker Br 23: 17291 ± 35 Da;
Marker Br 24: 17419 ± 35 Da;
Marker Br 25: 21103 ± 42 Da;
Marker Br 26: 40297 ± 81 Da;
Marker Br 27: 4488 ± 9 Da;
Marker Br 28: 4646 ± 9 Da;
Marker Br 29: 80428 ± 402 Da;
Marker Br 30: 11757 ± 24 Da;
Marker Br 31: 4487 ± 9 Da;
Marker Br 32: 5360 ± 11 Da;
Marker Br 33: 11773 ± 24 Da;
Marker Br 34: 13977 ± 28 Da;
Marker Br 35: 14004 ± 28 Da;
Marker Br 36: 51700 ± 259 Da; and
Marker Br 37: 153894 ± 769 Da;

wherein the method comprises detecting the marker or markers by an immunoassay.

40. A purified protein selected from:

Marker Br 1: 4170 ± 8 Da;
Marker Br 2: 21080 ± 42 Da;
Marker Br 3: 9339 ± 19 Da;
Marker Br 4: 28308 ± 57 Da;
Marker Br 5: 28344 ± 57 Da;
Marker Br 6: 4148 ± 9 Da;
Marker Br 7: 5634 ± 11 Da;
Marker Br 8: 6520 ± 13 Da;
Marker Br 9: 6955 ± 14 Da;
Marker Br 10: 7507 ± 15 Da;
Marker Br 11: 9116 ± 16 Da;
Marker Br 12: 9453 ± 19 Da;
Marker Br 13: 17310 ± 35 Da;
Marker Br 14: 89805 ± 449 Da;
Marker Br 15: 4256 ± 9 Da;
Marker Br 16: 4357 ± 9 Da;
Marker Br 17: 4470 ± 9 Da;
Marker Br 18: 9292 ± 19 Da;
Marker Br 19: 9335 ± 19 Da;
Marker Br 20: 11758 ± 24 Da;
Marker Br 21: 11776 ± 24 Da;
Marker Br 22: 13913 ± 28 Da;
Marker Br 23: 17291 ± 35 Da;
Marker Br 24: 17419 ± 35 Da;
Marker Br 25: 21103 ± 42 Da;
Marker Br 26: 40297 ± 81 Da;
Marker Br 27: 4488 ± 9 Da;
Marker Br 28: 4647 ± 9 Da;
Marker Br 29: 80428 ± 402 Da;
Marker Br 30: 11757 ± 24 Da;
Marker Br 31: 4487 ± 9 Da;
Marker Br 32: 5360 ± 11 Da;
Marker Br 33: 11773 ± 24 Da;
Marker Br 34: 13977 ± 28 Da;

Marker Br 35: 14004 ± 28 Da;

Marker Br 36: 51700 ± 259 Da; and

Marker Br 37: 153894 ± 769 Da.

41. A kit comprising:

(a) a substrate comprising an adsorbent attached thereto, wherein the adsorbent is capable of retaining at least one protein marker selected from:

Marker Br 1: 4170 ± 8 Da;

Marker Br 2: 21080 ± 42 Da;

Marker Br 3: 9339 ± 19 Da;

Marker Br 4: 28308 ± 57 Da;

Marker Br 5: 28344 ± 57 Da;

Marker Br 6: 4148 ± 9 Da;

Marker Br 7: 5634 ± 11 Da;

Marker Br 8: 6520 ± 13 Da;

Marker Br 9: 6955 ± 14 Da;

Marker Br 10: 7507 ± 15 Da;

Marker Br 11: 9116 ± 16 Da;

Marker Br 12: 9453 ± 19 Da;

Marker Br 13: 17310 ± 35 Da;

Marker Br 14: 89805 ± 449 Da;

Marker Br 15: 4256 ± 9 Da;

Marker Br 16: 4357 ± 9 Da;

Marker Br 17: 4470 ± 9 Da;

Marker Br 18: 9292 ± 19 Da;

Marker Br 19: 9335 ± 19 Da;

Marker Br 20: 11758 ± 24 Da;

Marker Br 21: 11776 ± 24 Da;

Marker Br 22: 13913 ± 28 Da;

Marker Br 23: 17291 ± 35 Da;

Marker Br 24: 17419 ± 35 Da;

Marker Br 25: 21103 ± 42 Da;

Marker Br 26: 40297 ± 81 Da;

Marker Br 27: 4488 ± 9 Da;

Marker Br 28: 4647 ± 9 Da;

Marker Br 29: 80428 ± 402 Da;

Marker Br 30: 11757 ± 24 Da;

Marker Br 31: 4487 ± 9 Da;

Marker Br 32: 5360 ± 11 Da;

Marker Br 33: 11773 ± 24 Da;

Marker Br 34: 13977 ± 28 Da;

Marker Br 35: 14004 ± 28 Da;

Marker Br 36: 51700 ± 259 Da; and

Marker Br 37: 153894 ± 769 Da; and

(b) instructions to detect the marker or markers by contacting a sample with the adsorbent and detecting the marker or markers retained by the adsorbent.

FIG. 1

Protein Profiling

Sequential Spin Columns and ProteinChip arrays

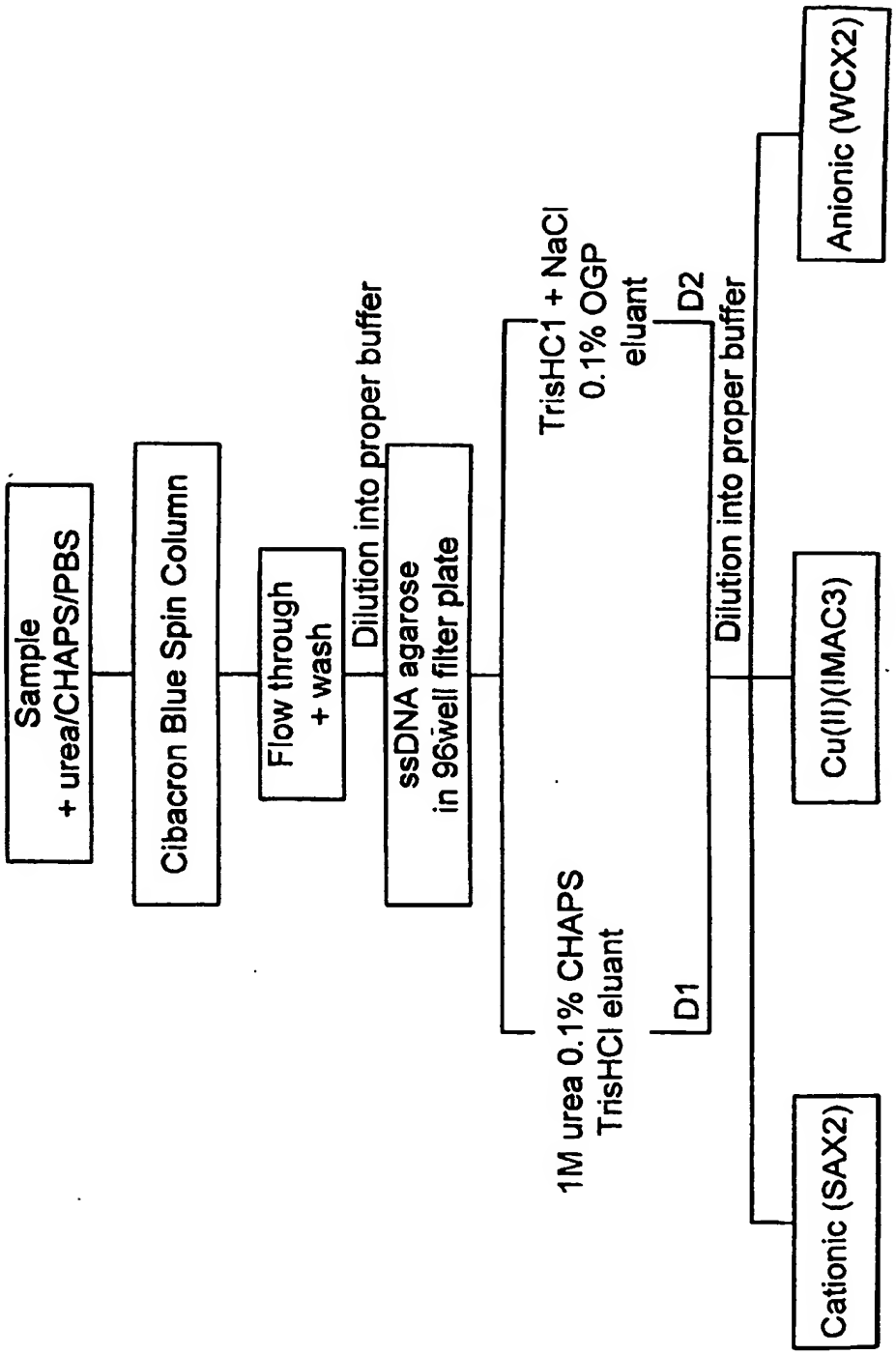


FIG. 2
Protein Profiling
Sequential Spin Columns and ProteinChip arrays

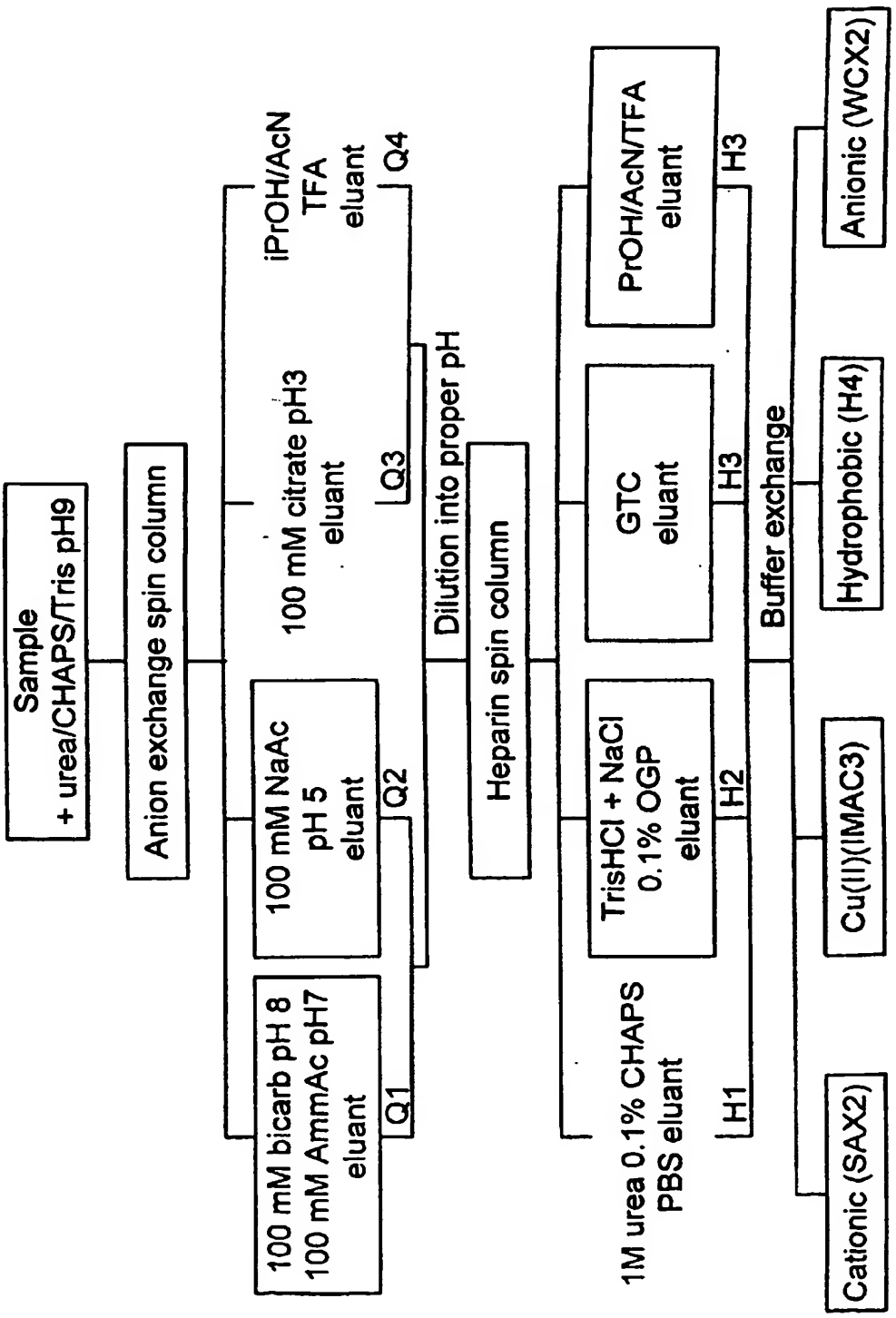


FIG. 3

Stacked Lectin Spin Column